

ON THE NUMBER OF SOLUTIONS IN RANDOM GRAPH  $k$ -COLOURING

FELICIA RASSMANN\*

**ABSTRACT.** Let  $k \geq 3$  be a fixed integer. We exactly determine the asymptotic distribution of  $\ln Z_k(G(n, m))$ , where  $Z_k(G(n, m))$  is the number of  $k$ -colourings of the random graph  $G(n, m)$ . A crucial observation to this aim is that the fluctuations in the number of colourings can be attributed to the fluctuations in the number of small cycles in  $G(n, m)$ . Our result holds for a wide range of average degrees, and for  $k$  exceeding a certain constant  $k_0$  it covers all average degrees up to the so-called *condensation phase transition*.

## 1. INTRODUCTION

**1.1. Background and motivation.** Going back to the ground-breaking paper of Erdős and Rényi [15] in 1960, the study of the random graph colouring problem has attained a lot of attention and innumerable articles have been published in this area of research over the years. In the most frequently studied model, a random graph  $G(n, m)$  on the vertex set  $[n] = \{1, \dots, n\}$  with precisely  $m$  edges is drawn uniformly at random from all such graphs.

A question that has turned out to be a very challenging one is how to choose  $n$  and  $m$  to obtain a random graph that is colourable w.h.p.. Or, put differently, whether a random graph with given  $n$  and  $m$  can be coloured with a fixed number of colours, thus determining its chromatic number.

Beginning in the 1990s, considerable progress has been made in the case of *sparse* random graphs, where  $m = O(n)$  as  $n \rightarrow \infty$ . Much effort has been devoted to studying the typical value of the chromatic number of  $G(n, m)$  [3, 9, 19, 21] and its concentration [4, 20, 29]. Several experiments and simulations led to the hypothesis, that, when changing the ratio of edges to variables, there is a transition from a regime where the random graph is colourable w.h.p. to the one where it is not w.h.p.. Furthermore, the observation was that this transition does not happen smoothly, suggesting the existence of a sharp satisfiability threshold. Indeed, in 1999, Achlioptas and Friedgut [2] proved the existence of a *sharp threshold sequence*  $d_{k,\text{col}}(n)$  for any  $k \geq 3$ , meaning that for any fixed  $\varepsilon > 0$  the random graph  $G(n, m)$  is  $k$ -colourable w.h.p. if  $2m/n < d_{k,\text{col}}(n) - \varepsilon$ , whereas  $G(n, m)$  fails to be  $k$ -colourable w.h.p. if  $2m/n > d_{k,\text{col}}(n) + \varepsilon$ . This threshold sequence is non-uniform, i.e. it is a function of  $n$  and although it is broadly believed to converge for  $n$  tending to infinity, this has not been established up to now. Also, in spite of continued efforts, the exact value of this threshold remains unknown up to date. The best current bounds [11, 13] on  $d_{k,\text{col}}(n)$  show that there is a sequence  $(\gamma_k)_{k \geq 3}$ ,  $\lim_{k \rightarrow \infty} \gamma_k = 0$ , such that

$$(2k - 1) \ln k - 2 \ln 2 - \gamma_k \leq \liminf_{n \rightarrow \infty} d_{k,\text{col}}(n) \leq \limsup_{n \rightarrow \infty} d_{k,\text{col}}(n) \leq (2k - 1) \ln k - 1 + \gamma_k.$$

Yet, there exist predictions by statistical physicists regarding the precise location of this threshold. They developed a method called *cavity method* that allowed them to gain insights into the combinatorial structure of the random graph colouring problem and to understand the significance of *typical*  $k$ -colourings, i.e.  $k$ -colourings chosen uniformly at random from the set of all  $k$ -colourings, on both the combinatorial and algorithmic aspects of the problem [18]. What is more, this method has also been used to predict a further phase transition shortly before the colouring threshold. This transition  $d_{k,\text{cond}}$  has been named *condensation* and its existence and location have rigorously been determined in 2014 by Bapst et al. [7]. Under the assumption that  $k \geq k_0$  for a certain constant  $k_0$  it is possible to calculate the number  $d_{k,\text{cond}}$  precisely [7], and an asymptotic expansion in  $k$  yields

$$d_{k,\text{cond}} = (2k - 1) \ln k - 2 \ln 2 + \gamma_k, \quad \text{where } \lim_{k \rightarrow \infty} \gamma_k = 0.$$

The condensation transition plays a very important role for several reasons. It marks the point where the behaviour of the number of solutions changes significantly, as does the geometry of the solution space [1, 22]. The prediction

---

*Date:* September 15, 2016.

\*The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 278857-PTCC.

states that while two  $k$ -colourings chosen uniformly at random tend to be uncorrelated before the condensation threshold, they typically exhibit long-range correlations afterwards [23]. Furthermore, the condensation transition persists, in contrast to the colourability transition, also for finite inverse temperatures [8]. In recent work, it has been proved that the condensation transition is also related to the information theoretic threshold in the stochastic block model [5, 25], where it marks the point from which on it is possible to decide whether a random graph has been drawn from a planted distribution or not.

By obtaining an exact expression for the asymptotic distribution of the logarithm of the number of solutions up to the condensation threshold  $d_{k,\text{cond}}$ , in the present paper we give a definite and complete answer to the question about the relationship between the planted model and the Gibbs distribution. Furthermore, we show that the fluctuations in the number of solutions can completely be attributed to the presence of short cycles, thereby eliminating the possibility of other influencing factors.

For a graph  $G$  on  $n$  vertices, we let  $Z_k(G)$  be the number of  $k$ -colourings (also called *solutions*) of  $G$ , which are maps  $\sigma : [n] \rightarrow [k]$  such that  $\sigma(i) \neq \sigma(j)$  for all edges  $\{i, j\}$  of  $G$ . We always consider *sparse* random graphs  $G(n, m)$  where  $m = O(n)$ . As we are going to need a very precise computation of the first and second moment of the number of  $k$ -colourings of  $G(n, m)$ , we distinguish the parameter  $d'$ , which is such that  $m = \lceil d'n/2 \rceil$ , from  $d = 2m/n$ , which arises naturally in the computations of the first and second moment. We note that  $d' \sim d$ , although  $d = d(n)$  might vary with  $n$ , whereas  $d'$  is assumed to be fixed as  $n \rightarrow \infty$ .

**1.2. Results.** We show that under certain conditions the number  $Z_k(G(n, m))$  of  $k$ -colourings of the random graph is concentrated tightly and determine the distribution of  $\ln Z_k(G(n, m)) - \ln \mathbb{E}[Z_k(G(n, m))]$  asymptotically in a density regime up to the condensation transition.

Before we state the result, we introduce the following notation. For  $k \geq 3$ , we define

$$d_{k,\text{cond}} = \sup \left\{ d' > 0 : \liminf_{n \rightarrow \infty} \mathbb{E} \left[ Z_k(G(n, m))^{1/n} \right] = k(1 - 1/k)^{d'/2} \right\}. \quad (1.1)$$

This definition is motivated by the well-known fact that

$$\mathbb{E} [Z_k(G(n, m))] = \Theta(k^n (1 - 1/k)^m).$$

Jensen's inequality shows that  $\limsup_{n \rightarrow \infty} \mathbb{E} [Z_k(G(n, m))^{1/n}] \leq k(1 - 1/k)^{d'/2}$  for all  $d'$  and this upper bound is tight up to the density  $d_{k,\text{cond}}$ .

**Theorem 1.1.** *There is a constant  $k_0 > 3$  such that the following is true. Assume either that  $k \geq 3$  and  $d' \leq 2(k-1) \ln(k-1)$  or that  $k \geq k_0$  and  $d' < d_{k,\text{cond}}$ . Further, let*

$$\lambda_l = \frac{d^l}{2l} \quad \text{and} \quad \delta_l = \frac{(-1)^l}{(k-1)^{l-1}}$$

for  $l \geq 2$ . Let  $(X_l)_l$  be a family of independent Poisson variables with  $\mathbb{E}[X_l] = \lambda_l$ , all defined on the same probability space. Then the random variable

$$W = \sum_{l \geq 3} X_l \ln(1 + \delta_l) - \lambda_l \delta_l - d^2/(4(k-1))$$

satisfies  $\mathbb{E}|W| < \infty$  and  $\ln Z_k(G(n, m)) - \ln \mathbb{E}[Z_k(G(n, m))]$  converges in distribution to  $W$ .

**Remark 1.2.** By definition,  $W$  has an infinitely divisible distribution. It was shown in [16] that the random variable  $W' = \exp[W]$  converges almost surely and in  $L^2$  with  $\mathbb{E}[W'] = 1$  and  $\mathbb{E}[W'^2] = \exp[\sum_l \lambda_l \delta_l^2]$ . Thus, by Jensen's inequality it follows that  $\mathbb{E}[W] \leq 0$  and  $\mathbb{E}[W^2] \leq \sum_l \lambda_l \delta_l^2$ .

**1.3. Discussion and further related work.** The crucial observation that the proof of Theorem 1.1 builds upon is that the fluctuations of  $\ln Z_k(G(n, m))$  can be attributed to variations in the number of cycles of bounded length in the random graph. This has first been observed in [6] and has been used to determine the order of magnitude of the fluctuations of  $\ln Z_k(G(n, m))$  in the random graph colouring problem. Following this result, the asymptotic distribution of the logarithm of the number of solutions has been established for random regular  $k$ -SAT [14] and random hypergraph 2-colouring [26]. Our result Theorem 1.1 refines the analysis from [14] to obtain this asymptotic distribution also in random graph  $k$ -colouring in a broad density regime up to the condensation transition  $d_{k,\text{cond}}$  (for large values of  $k$ ). The proof combines the second moment arguments from Achlioptas and Naor [3] and its enhancements from [7, 13] with the “small subgraph conditioning”. This method was originally developed in [27, 28]

and extended by Janson [16] to obtain limiting distributions. It has been frequently used in random regular graph problems (see [30] for an enlightening survey), e.g. in [17] and [12] to upper-bound the chromatic number of the random  $d$ -regular graph, as the sharp threshold result [2] does not apply for this problem. More recently, it has also been used to obtain results in the stochastic block model [25] and to determine the satisfiability threshold for positive 1-in- $k$ -SAT [24]. Unfortunately, Janson's result does not apply directly in our case and instead we have to perform a variance analysis along the lines of [28], very analogue to [14, 26]. The reason for this is that in contrast to [6], where only bounds on the fluctuation of  $\ln Z_k$  were proven, we aim at a statement about its asymptotic *distribution*. Thus, in the present paper it does not suffice to consider colourings with *balanced* colour classes (with a deviation of  $o(n^{-1/2})$  from their typical value), but we have to get a handle on all colourings providing a positive contribution. To this aim, we collect together colourings exhibiting similar colour class sizes. This results in the need to not only consider one random variable, but a growing number of random variables simultaneously. We expect that it is possible to apply a combination of the second moment method and small subgraph conditioning to a variety of further random constraint problems, such as e.g. random  $k$ -NAESAT, random  $k$ -XORSAT or random hypergraph  $k$ -colourability. However, for asymmetric problems like the well-known benchmark problem random  $k$ -SAT, we expect that the logarithm of the number of satisfying assignments exhibits stronger fluctuations and we doubt that a result similar to ours can be established.

**1.4. Preliminaries and notation.** We always assume that  $n \geq n_0$  is large enough for our various estimates to hold and denote by  $[n]$  the set  $\{1, \dots, n\}$ .

We use the standard  $O$ -notation when referring to the limit  $n \rightarrow \infty$ . Thus,  $f(n) = O(g(n))$  means that there exist  $C > 0$ ,  $n_0 > 0$  such that for all  $n > n_0$  we have  $|f(n)| \leq C \cdot |g(n)|$ . In addition, we use the standard symbols  $o(\cdot)$ ,  $\Omega(\cdot)$ ,  $\Theta(\cdot)$ . In particular,  $o(1)$  stands for a term that tends to 0 as  $n \rightarrow \infty$ . Furthermore, the notation  $f(n) \sim g(n)$  means that  $f(n) = g(n)(1 + o(1))$  or equivalently  $\lim_{n \rightarrow \infty} f(n)/g(n) = 1$ . Besides taking the limit  $n \rightarrow \infty$ , at some point we need to consider the limit  $\nu \rightarrow \infty$  for some number  $\nu \in \mathbb{N}$ . Thus, we introduce  $f(n, \nu) \sim_\nu g(n, \nu)$  meaning that  $\lim_{\nu \rightarrow \infty} \lim_{n \rightarrow \infty} f(n, \nu)/g(n, \nu) = 1$ .

If  $p = (p_1, \dots, p_l)$  is a vector with entries  $p_i \geq 0$ , then we let

$$\mathcal{H}(p) = - \sum_{i=1}^l p_i \ln p_i.$$

Here and throughout, we use the convention that  $0 \ln 0 = 0$ . Hence, if  $\sum_{i=1}^l p_i = 1$ , then  $\mathcal{H}(p)$  is the entropy of the probability distribution  $p$ . Further, for a number  $x$  and an integer  $h > 0$  we let  $(x)_h = x(x-1) \cdots (x-h+1)$  denote the  $h$ th falling factorial of  $x$ .

For the sake of simplicity, we choose to prove Theorem 1.1 using the random graph model  $\mathcal{G}(n, m)$ . This is a random (multi-)graph on the vertex set  $[n]$  obtained by choosing  $m$  edges  $e_1, \dots, e_m$  of the complete graph on  $n$  vertices uniformly and independently at random (i.e., with replacement). In this model we may choose the same edge more than once. However, the following statement shows that for sparse random graphs the probability of this event is bounded away from 1:

**Fact 1.3.** *Assume that  $m = m(n)$  is a sequence such that  $m = O(n)$  and let  $\mathcal{A}_n$  be the event that  $\mathcal{G}(n, m)$  has no multiple edges. Then there is a constant  $c > 0$  such that  $\lim_{n \rightarrow \infty} \mathbb{P}[\mathcal{A}_n] > c$ .*

## 2. OUTLINE OF THE PROOF

To determine bounds on  $Z_k(\mathcal{G}(n, m))$ , it will be necessary to control the size of the colour classes. To formalize this, we introduce the following notation. For a map  $\sigma : [n] \rightarrow [k]$ , we define

$$\rho(\sigma) = (\rho_1(\sigma), \dots, \rho_k(\sigma)), \quad \text{where } \rho_i(\sigma) = |\sigma^{-1}(i)|/n \quad \text{for } i = 1 \dots k.$$

Thus,  $\rho(\sigma)$  is a probability distribution on  $[k]$ , to which we refer as the *colour density* of  $\sigma$ .

Let  $\mathcal{A}_k(n)$  signify the set of all possible colour densities  $\rho(\sigma)$  for  $\sigma : [n] \rightarrow [k]$ . Further, let  $\mathcal{A}_k$  be the set of all probability distributions  $\rho = (\rho_1, \dots, \rho_k)$  on  $[k]$ , and let  $\rho^* = (1/k, \dots, 1/k)$  signify the barycentre of  $\mathcal{A}_k$ .

In order to simplify the notation, for the rest of the paper we assume that  $\omega, \nu$  are odd natural numbers, formally we define  $N = \{2i - 1 : i \in \mathbb{N}\}$  and let  $\omega, \nu \in N$ . We say that  $\rho = (\rho_1, \dots, \rho_k) \in \mathcal{A}_k(n)$  is  $(\omega, \nu)$ -balanced if

$$\rho_i \in \left[ \frac{1}{k} - \frac{\omega}{\sqrt{n}}, \frac{1}{k} + \frac{\omega}{\sqrt{n}} \right) \quad \text{for all } i \in [k]$$

and let  $\mathcal{A}_{k,\omega}(n)$  denote the set of all  $(\omega, n)$ -balanced  $\rho \in \mathcal{A}_k(n)$ . As we will see, in order to prove statements about the number  $Z_k$  of all solutions, it suffices to consider solutions  $\sigma$  with  $\rho(\sigma) \in \mathcal{A}_{k,\omega}(n)$ . We let  $Z_{k,\omega}(G)$  signify the number of  $(\omega, n)$ -balanced  $k$ -colourings of a graph  $G$  on  $[n]$ , i.e.  $k$ -colourings  $\sigma$  such that  $\rho(\sigma) \in \mathcal{A}_{k,\omega}(n)$ .

Since verifying the required properties to apply small subgraph conditioning directly for the random variable  $Z_{k,\omega}$  is very intricate, we break  $Z_{k,\omega}$  down into smaller contributions, for which we determine the first and second moment in the following sections.

To this aim, we decompose the set  $\mathcal{A}_{k,\omega}(n)$  into smaller sets. We define

$$S_{k,\omega,\nu} = \left\{ s \in \mathbb{Z}^k : \|s\|_1 = 2i, i \in \mathbb{N}, i \leq \frac{\omega\nu - 1}{2} \right\}. \quad (2.1)$$

$S_{k,\omega,\nu}$  contains vectors that we use as centres of disjoint 'balls' to partition the set  $\mathcal{A}_{k,\omega}(n)$ : For  $s = (s_1, \dots, s_k) \in S_{k,\omega,\nu}$ , we let  $\rho^{k,\omega,\nu,s} \in \mathbb{R}^k$  be the vector with components

$$\rho_i^{k,\omega,\nu,s} = \frac{1}{k} + \frac{s_i}{\nu\sqrt{n}}. \quad (2.2)$$

Further, we let  $\mathcal{A}_{k,\omega,\nu}^s(n)$  be the set of all colour densities  $\rho \in \mathcal{A}_{k,\omega}(n)$  such that

$$\rho_i \in \left[ \rho_i^{k,\omega,\nu,s} - \frac{1}{\nu\sqrt{n}}, \rho_i^{k,\omega,\nu,s} + \frac{1}{\nu\sqrt{n}} \right).$$

For a graph  $G$ , we denote by  $Z_{k,\omega,\nu}^s(G)$  the number of 2-colourings  $\sigma$  such that  $\rho(\sigma) \in \mathcal{A}_{k,\omega,\nu}^s(n)$ . For each fixed  $\nu$ , we have  $Z_{k,\omega} = \sum_{s \in S_{k,\omega,\nu}} Z_{k,\omega,\nu}^s$  and our strategy is to apply small subgraph conditioning to the random variables  $Z_{k,\omega,\nu}^s$  rather than directly to  $Z_k$ . But first, we will calculate the first moments of  $Z_k$  and  $Z_{k,\omega}$  in Section 3 to obtain the following.

**Proposition 2.1.** *Fix an integer  $k \geq 3$  and a number  $d' \in (0, \infty)$ . Let  $\omega > 0$ . Then*

$$\mathbb{E}[Z_k(\mathcal{G}(n, m))] = \Theta(k^n (1 - 1/k)^m) \quad \text{and} \quad \lim_{\omega \rightarrow \infty} \liminf_{n \rightarrow \infty} \frac{\mathbb{E}[Z_{k,\omega}(\mathcal{G}(n, m))]}{\mathbb{E}[Z_k(\mathcal{G}(n, m))]} = 1.$$

As discussed in Section 1.3, the key observation the proof is based on is that the fluctuations of  $Z_k(\mathcal{G}(n, m))$  can be attributed to fluctuations in the number of cycles of a bounded length. Hence, for an integer  $l \geq 2$  we let  $C_{l,n}$  denote the number of cycles of length exactly  $l$  in  $\mathcal{G}(n, m)$ . Let

$$\lambda_l = \frac{d^l}{2l} \quad \text{and} \quad \delta_l = \frac{(-1)^l}{(k-1)^{l-1}}. \quad (2.3)$$

The following fact shows that  $C_{2,n}, \dots$  are asymptotically independent Poisson variables (e.g. [10, Theorem 5.16]):

**Fact 2.2.** *If  $c_2, \dots, c_L$  are non-negative integers, then*

$$\lim_{n \rightarrow \infty} \mathbb{P}[\forall 2 \leq l \leq L : C_{l,n} = c_l] = \prod_{l=2}^L \mathbb{P}[\text{Po}(\lambda_l) = c_l].$$

In Section A.3 the impact of the cycle counts  $C_{l,n}$  on the first moment of  $Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))$  is investigated. As this was already done in [6], we carry it out in the present work only for the sake of completeness. The result is the following:

**Proposition 2.3.** *Assume that  $k \geq 3$  and  $d' \in (0, \infty)$ . Then*

$$\sum_{l=2}^{\infty} \lambda_l \delta_l^2 < \infty.$$

Moreover, let  $\omega, \nu \in \mathbb{N}$  and  $c_2, \dots, c_L$  be non-negative integers. Then

$$\frac{\mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \forall 2 \leq l \leq L : C_{l,n} = c_l]}{\mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))]} \sim \prod_{l=2}^L [1 + \delta_l]^{c_l} \exp[-\delta_l \lambda_l]. \quad (2.4)$$

Additionally, to apply small subgraph conditioning, we have to determine the second moment of  $Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))$  very precisely. This step constitutes the main technical work of this paper. We consider two regimes of  $d'$  and  $k$  separately. In the simpler case, based on the second moment argument from [3], we obtain the following result.

**Proposition 2.4.** *Assume that  $k \geq 3$  and  $d' < 2(k-1) \ln(k-1)$ . Then*

$$\frac{\mathbb{E} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))^2 \right]}{\mathbb{E} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) \right]^2} \sim \exp \left[ \sum_{l \geq 2} \lambda_l \delta_l^2 \right].$$

The second regime of  $d'$  and  $k$  is that  $k \geq k_0$  for a certain constant  $k_0 \geq 3$  and  $d' < d_{k,\text{cond}}$  (with  $d_{k,\text{cond}}$  the number defined in (1.1)). In this case, we replace  $Z_{k,\omega,\nu}^s$  by the slightly tweaked random variable  $\tilde{Z}_{k,\omega,\nu}^s$  used in the second moment arguments from [7, 13].

**Proposition 2.5.** *There is a constant  $k_0 \geq 3$  such that the following is true. Assume that  $k \geq k_0$  and  $2(k-1) \ln(k-1) \leq d' < d_{k,\text{cond}}$ . Then for each  $\omega, \nu \in N$  and  $s \in S_{k,\omega,\nu}$  there exists an integer-valued random variable  $0 \leq \tilde{Z}_{k,\omega,\nu}^s \leq Z_{k,\omega,\nu}^s$  such that*

$$\begin{aligned} \mathbb{E} \left[ \tilde{Z}_{k,\omega,\nu}^s(\mathcal{G}(n, m)) \right] &\sim \mathbb{E} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) \right] \quad \text{and} \\ \frac{\mathbb{E} \left[ \tilde{Z}_{k,\omega,\nu}^s(\mathcal{G}(n, m))^2 \right]}{\mathbb{E} \left[ \tilde{Z}_{k,\omega,\nu}^s(\mathcal{G}(n, m)) \right]^2} &\leq (1 + o(1)) \exp \left[ \sum_{l \geq 2} \lambda_l \delta_l^2 \right]. \end{aligned} \tag{2.5}$$

The proofs of Propositions 2.4 and 2.5 appear at the end of Section 4. In order to apply small subgraph conditioning to the random variable  $\tilde{Z}_{k,\omega,\nu}^s$ , we need to investigate the impact of  $C_{l,n}$  on the first moment of  $\tilde{Z}_{k,\omega,\nu}^s$ . Thus, we need a similar result as Proposition 2.3 for  $\tilde{Z}_{k,\omega,\nu}^s$ . Fortunately, instead of having to reiterate the proof of Proposition 2.3, we obtain the following by combining Proposition 2.3 with (2.5):

**Corollary 2.6.** *Let  $c_2, \dots, c_L$  be non-negative integers. With the assumptions and notation of Proposition 2.5 we have*

$$\frac{\mathbb{E} \left[ \tilde{Z}_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \forall 2 \leq l \leq L : C_{l,n} = c_l \right]}{\mathbb{E} \left[ \tilde{Z}_{k,\omega,\nu}^s(\mathcal{G}(n, m)) \right]} \sim \prod_{l=2}^L [1 + \delta_l]^{c_l} \exp [-\delta_l \lambda_l].$$

As the proof is nearly identical to the one in [6], we defer it to Appendix A.

The aim is now to derive Theorem 1.1 from Propositions 2.1-2.4. The key observation is that the variance of the random variables  $Z_{k,\omega,\nu}^s$  is affected by the presence of cycles of bounded length and that this is the only significant influence. As a consequence, conditioning on the small cycle counts up to some preselected length reduces the variance of  $Z_{k,\omega,\nu}^s$ . What is maybe surprising is that conditioning on the number of enough small cycles reduces the variance to any desired fraction of  $\mathbb{E}[Z_{k,\omega,\nu}^s]^2$ .

As done in [14, 26], the arguments we use are similar to the small subgraph conditioning from [16, 28]. But we do not refer to any technical statements from [16, 28] directly because instead of working only with the random variable  $Z_k$  we need to control all  $Z_{k,\omega,\nu}^s$  for fixed  $\omega, \nu \in N$  simultaneously. In fact, ultimately we have to take  $\nu \rightarrow \infty$  and  $\omega \rightarrow \infty$  as well. Our line of argument follows the path beaten in [14, 26] and the following three lemmas are nearly identical to the ones derived there.

For  $L > 2$ , let  $\mathcal{F}_L = \mathcal{F}_{L,n}(d, k)$  be the  $\sigma$ -algebra generated by the random variables  $C_{l,n}$  with  $2 \leq l \leq L$ . The set of all graphs can be divided into groups according to the small cycle counts: For each  $L \geq 2$ , the decomposition of the variance of  $Z_{k,\omega,\nu}^s$  yields

$$\text{Var} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) \right] = \text{Var} \left[ \mathbb{E} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \mathcal{F}_L \right] \right] + \mathbb{E} \left[ \text{Var} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \mathcal{F}_L \right] \right],$$

meaning that the variance can be written as the variance of the group mean plus the expected value of the variance within a group. The term  $\text{Var} \left[ \mathbb{E} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \mathcal{F}_L \right] \right]$  accounts for the amount of variance induced by the fluctuations of the number of cycles of length at most  $L$ . The strategy when using small subgraph conditioning is to bound the second summand, which is the expected conditional variance

$$\mathbb{E} \left[ \text{Var} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \mathcal{F}_L \right] \right] = \mathbb{E} \left[ \mathbb{E} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))^2 | \mathcal{F}_L \right] - \mathbb{E} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \mathcal{F}_L \right]^2 \right].$$

In the following lemma we show that in fact in the limit of large  $L$  and  $n$  this quantity is negligible. This implies that conditioned on the number of short cycles the variance vanishes and thus the limiting distribution of  $\ln Z_{k,\omega,\nu}^s$  is just the limit of  $\ln \mathbb{E} [Z_{k,\omega,\nu}^s | \mathcal{F}_L]$  as  $n, L \rightarrow \infty$ . This limit is determined by the joint distribution of the number of short cycles.

**Lemma 2.7.** *Let  $k \geq 3$  and  $d' \in (0, \infty)$ . For any  $\omega, \nu \in N$  and  $s \in S_{k,\omega,\nu}$ , we have*

$$\limsup_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} \left[ \frac{\mathbb{E} [Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))^2 | \mathcal{F}_L] - \mathbb{E} [Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \mathcal{F}_L]^2}{\mathbb{E} [Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))]^2} \right] = 0.$$

*Proof.* Fix  $\omega, \nu \in N$  and set  $Z_s = Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))$ . Using Fact 2.2 and equation (2.4) from Proposition 2.3 we can choose for any  $\varepsilon > 0$  a constant  $B = B(\varepsilon)$  and  $L \geq L_0(\varepsilon)$  large enough such that for each large enough  $n \geq n_0(\varepsilon, B, L)$  we have for any  $s \in S_{k,\omega,\nu}$ :

$$\begin{aligned} \mathbb{E} [\mathbb{E} [Z_s | \mathcal{F}_L]^2] &\geq \sum_{c_1, \dots, c_L \leq B} \mathbb{E} [Z_s | \forall 2 \leq l \leq L : C_{l,n} = c_l]^2 \mathbb{P} [\forall 2 \leq l \leq L : C_{l,n} = c_l] \\ &\geq \exp [-\varepsilon] \mathbb{E} [Z_s]^2 \sum_{c_1, \dots, c_L \leq B} \prod_{l=2}^L [(1 + \delta_l)^{c_l} \exp [-\lambda_l \delta_l]]^2 \mathbb{P} [\text{Po}(\lambda_l) = c_l] \\ &= \exp [-\varepsilon] \mathbb{E} [Z_s]^2 \sum_{c_1, \dots, c_L \leq B} \prod_{l=2}^L \frac{[(1 + \delta_l)^2 \lambda_l]^{c_l}}{c_l! \exp [2\lambda_l \delta_l + \lambda_l]} \\ &\geq \mathbb{E} [Z_s]^2 \exp \left[ -2\varepsilon + \sum_{l=2}^L \delta_l^2 \lambda_l \right]. \end{aligned} \tag{2.6}$$

The tower property for conditional expectations and the standard formula for the decomposition of the variance yields

$$\mathbb{E} [Z_s^2] = \mathbb{E} [\mathbb{E} [Z_s^2 | \mathcal{F}_L]] = \mathbb{E} [\mathbb{E} [Z_s^2 | \mathcal{F}_L] - \mathbb{E} [Z_s | \mathcal{F}_L]^2] + \mathbb{E} [\mathbb{E} [Z_s | \mathcal{F}_L]^2]$$

and thus, using (2.6) we have

$$\frac{\mathbb{E} [\mathbb{E} [Z_s^2 | \mathcal{F}_L] - \mathbb{E} [Z_s | \mathcal{F}_L]^2]}{\mathbb{E} [Z_s]^2} \leq \frac{\mathbb{E} [Z_s^2]}{\mathbb{E} [Z_s]^2} - \exp \left[ -2\varepsilon + \sum_{l=2}^L \delta_l^2 \lambda_l \right]. \tag{2.7}$$

Finally, the estimate  $\exp[-x] \geq 1 - x$  for  $|x| < 1/8$  combined with (2.7) and Proposition 2.4 implies that for large enough  $\nu, n, L$  and each  $s \in S_{k,\omega,\nu}$  we have

$$\frac{\mathbb{E} [\mathbb{E} [Z_s^2 | \mathcal{F}_L] - \mathbb{E} [Z_s | \mathcal{F}_L]^2]}{\mathbb{E} [Z_s]^2} \leq 2\varepsilon \exp \left[ \sum_{l=2}^{\infty} \delta_l^2 \lambda_l \right].$$

As this holds for any  $\varepsilon > 0$  and by equation (2.3) the expression  $\exp [\sum_{l=2}^{\infty} \delta_l^2 \lambda_l]$  is bounded, the proof of the lemma is completed by first taking  $n \rightarrow \infty$  and then  $L \rightarrow \infty$ .  $\square$

**Lemma 2.8.** *For any  $\alpha > 0$ , we have*

$$\limsup_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} [|Z_k(\mathcal{G}(n, m)) - \mathbb{E} [Z_k(\mathcal{G}(n, m)) | \mathcal{F}_L]| > \alpha \mathbb{E} [Z_k(\mathcal{G}(n, m))]] = 0.$$

*Proof.* To unclutter the notation, we set  $Z_k = Z_k(\mathcal{G}(n, m))$  and  $Z_{k,\omega} = Z_{k,\omega}(\mathcal{G}(n, m))$ . First we observe that Proposition 2.1 implies that for any  $\alpha > 0$  we can choose  $\omega \in N$  large enough such that

$$\liminf_{n \rightarrow \infty} \mathbb{E} [Z_{k,\omega}] > (1 - \alpha^2) \mathbb{E} [Z_k]. \tag{2.8}$$

We let  $\nu \in N$ . To prove the statement, we need to get a handle on the cases where the variables  $Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))$  deviate strongly from their conditional expectation  $\mathbb{E} [Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \mathcal{F}_L]$ . We let  $Z_s = Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))$  and



define

$$X_s = |Z_s - \mathbb{E}[Z_s|\mathcal{F}_L]| \cdot \mathbf{1}_{\{|Z_s - \mathbb{E}[Z_s|\mathcal{F}_L]| > \alpha \mathbb{E}[Z_s]\}}$$

and  $X = \sum_{s \in S_{k,\omega,\nu}} X_s$ . Then these definitions directly yield

$$\mathbb{P}[X < \alpha \mathbb{E}[Z_{k,\omega}]] \leq \mathbb{P}[|Z_{k,\omega} - \mathbb{E}[Z_{k,\omega}|\mathcal{F}_L]| < 2\alpha \mathbb{E}[Z_{k,\omega}]]. \quad (2.9)$$

By the definition of the  $X_s$ 's and Chebyshev's inequality it is true for every  $s$  that

$$\mathbb{E}[X_s|\mathcal{F}_L] \leq \sum_{j \geq 0} 2^{j+1} \alpha \mathbb{E}[Z_s] \mathbb{P}[|Z_s - \mathbb{E}[Z_s|\mathcal{F}_L]| > 2^j \alpha \mathbb{E}[Z_s]] \leq \frac{4 \text{Var}[Z_s|\mathcal{F}_L]}{\alpha \mathbb{E}[Z_s]}.$$

Hence, using that with Proposition 2.1 there is a number  $\beta = \beta(\alpha, \omega)$  such that  $\mathbb{E}[Z_s]/\mathbb{E}[Z_k] \leq \beta/(|S_{k,\omega,\nu}|)$  for all  $s \in S_{k,\omega,\nu}$  and  $n$  large enough, we have

$$\mathbb{E}[X|\mathcal{F}_L] \leq \sum_{s \in S_{k,\omega,\nu}} \frac{4 \text{Var}[Z_s|\mathcal{F}_L]}{\alpha \mathbb{E}[Z_s]} \leq \frac{4\beta \mathbb{E}[Z_k]}{\alpha |S_{k,\omega,\nu}|} \sum_{s \in S_{k,\omega,\nu}} \frac{\text{Var}[Z_s|\mathcal{F}_L]}{\mathbb{E}[Z_s]^2}.$$

Taking expectations, choosing  $\varepsilon = \varepsilon(\alpha, \beta, \omega)$  small enough and applying Lemma 2.7, we obtain

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_L]] \leq \frac{4\beta \mathbb{E}[Z_k]}{\alpha |S_{k,\omega,\nu}|} \sum_{s \in S_{k,\omega,\nu}} \frac{\mathbb{E}[\text{Var}[Z_s|\mathcal{F}_L]]}{\mathbb{E}[Z_s]^2} \leq \frac{4\beta \varepsilon \mathbb{E}[Z_k]}{\alpha} \leq \alpha^2 \mathbb{E}[Z_k]. \quad (2.10)$$

Using (2.9), Markov's inequality, (2.10) and (2.8), it follows that

$$\mathbb{P}[|Z_{k,\omega} - \mathbb{E}[Z_{k,\omega}|\mathcal{F}_L]| < 2\alpha \mathbb{E}[Z_{k,\omega}]] \geq 1 - 2\alpha. \quad (2.11)$$

Finally, the triangle inequality combined with Markov's inequality and equations (2.8) and (2.11) yields

$$\begin{aligned} \mathbb{P}[|Z_k - \mathbb{E}[Z_k|\mathcal{F}_L]| > \alpha \mathbb{E}[Z_k]] &\leq \mathbb{P}[|Z_k - Z_{k,\omega}| + |Z_{k,\omega} - \mathbb{E}[Z_{k,\omega}|\mathcal{F}_L]| + |\mathbb{E}[Z_{k,\omega}|\mathcal{F}_L] - \mathbb{E}[Z_k|\mathcal{F}_L]| > \alpha \mathbb{E}[Z_k]] \\ &\leq 3\alpha + \alpha/3 + 3\alpha < 7\alpha, \end{aligned}$$

which proves the statement.  $\square$

**Lemma 2.9.** *Let*

$$U_L = \sum_{l=2}^L C_{l,n} \ln(1 + \delta_l) - \lambda_l \delta_l. \quad (2.12)$$

Then  $\limsup_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E}[|U_L|] < \infty$  and further for any  $\varepsilon > 0$  we have

$$\limsup_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}[|\ln \mathbb{E}[Z_k(\mathcal{G}(n, m))|\mathcal{F}_L] - \ln \mathbb{E}[Z_k(\mathcal{G}(n, m))]| - U_L| > \varepsilon] = 0 \quad (2.13)$$

*Proof.* In a first step we show that  $\mathbb{E}[|U_L|]$  is uniformly bounded. As  $x - x^2 \leq \ln(1 + x) \leq x$  for  $|x| \leq 1/8$  we have for every  $l \leq L$ :

$$\mathbb{E}[|C_{l,n} \ln(1 + \delta_l) - \lambda_l \delta_l|] \leq \delta_l \mathbb{E}[|C_{l,n} - \lambda_l|] + \delta_l^2 \mathbb{E}[C_{l,n}].$$

Therefore, Fact 2.2 implies that

$$\mathbb{E}[|U_L|] \leq \sum_{l=2}^L \delta_l \sqrt{\lambda_l} + \delta_l^2 \lambda_l. \quad (2.14)$$

Proposition 2.3 ensures that  $\sum_l \delta_l^2 \lambda_l < \infty$ . Furthermore, as  $d' \leq (2k-1) \ln k$ , we have

$$\sum_l \delta_l \sqrt{\lambda_l} \leq \sum_l k^l 2^{-(k-1)l/2} < \infty$$

and thus (2.14) shows that  $\mathbb{E}[|U_L|]$  is uniformly bounded.

To prove (2.13), for given  $n$  and a constant  $B > 0$  we let  $C_B$  be the event that  $C_{l,n} < B$  for all  $l \leq L$ . Referring to Fact 2.2, we can find for each  $L, \varepsilon > 0$  a  $B > 0$  such that

$$\mathbb{P}[C_B] > 1 - \varepsilon. \quad (2.15)$$

To simplify the notation we set  $Z_k = Z_k(\mathcal{G}(n, m))$  and  $Z_{k,\omega} = Z_{k,\omega}(\mathcal{G}(n, m))$ . By Proposition 2.1 we can choose for any  $\alpha > 0$  a  $\omega > 0$  large enough such that  $\mathbb{E}[Z_{k,\omega}] > (1 - \alpha)\mathbb{E}[Z_k]$  for large enough  $n$ . Then Propositions 2.1 and 2.3 combined with Fact 2.2 imply that for any  $c_1, \dots, c_L \leq B$  and small enough  $\alpha = \alpha(\varepsilon, L, B)$  we have for  $n$  large enough:

$$\begin{aligned} \mathbb{E}[Z_k | \forall 2 \leq l \leq L : C_{l,n} = c_l] &\geq \mathbb{E}[Z_{k,\omega} | \forall 2 \leq l \leq L : C_{l,n} = c_l] \\ &\geq \exp[-\varepsilon] \mathbb{E}[Z_k] \prod_{l=2}^L (1 + \delta_l)^{c_l} \exp[-\delta_l \lambda_l]. \end{aligned} \quad (2.16)$$

On the other hand, for  $\alpha$  sufficiently small and large enough  $n$  we have

$$\begin{aligned} \mathbb{E}[Z_k | \forall 2 \leq l \leq L : C_{l,n} = c_l] &= \mathbb{E}[Z_k - Z_{k,\omega} | \forall 2 \leq l \leq L : C_{l,n} = c_l] + \mathbb{E}[Z_{k,\omega} | \forall 2 \leq l \leq L : C_{l,n} = c_l] \\ &\leq \frac{2\alpha \mathbb{E}[Z_k]}{\prod_{l=2}^L \mathbb{P}[\text{Po}(\lambda_l) = c_l]} + \mathbb{E}[Z_{k,\omega} | \forall 2 \leq l \leq L : C_{l,n} = c_l] \\ &\leq \exp[\varepsilon] \mathbb{E}[Z_k] \prod_{l=2}^L (1 + \delta_l)^{c_l} \exp[-\delta_l \lambda_l] \end{aligned} \quad (2.17)$$

Thus, the proof of (2.13) is completed by combining (2.15), (2.16), (2.17) and taking logarithms.  $\square$

*Proof of Theorem 1.1.* For  $L \geq 2$ , we define

$$W_L = \sum_{l=2}^L X_l \ln(1 + \delta_l) - \lambda_l \delta_l \quad \text{and} \quad W' = \sum_{l \geq 2} X_l \ln(1 + \delta_l) - \lambda_l \delta_l.$$

Then Fact 2.2 implies that for each  $L$  the random variables  $U_L$  defined in (2.12) converge in distribution to  $W_L$  as  $n \rightarrow \infty$ . Furthermore, because  $\sum_l \delta_l \sqrt{\lambda_l}, \sum_l \delta_l^2 \lambda_l < \infty$ , the martingale convergence theorem implies that  $W'$  is well-defined and that the  $W_L$  converge to  $W'$  almost surely as  $L \rightarrow \infty$ . Hence, from Lemmas 2.9 and 2.8 it follows that  $\ln Z_k(\mathcal{G}(n, m)) - \ln \mathbb{E}[Z_k(\mathcal{G}(n, m))]$  converges to  $W'$  in distribution, meaning that for any  $\varepsilon > 0$  we have

$$\lim_{n \rightarrow \infty} \mathbb{P}[|\ln Z_k(\mathcal{G}(n, m)) - \ln \mathbb{E}[Z_k(\mathcal{G}(n, m))]| - W'| > \varepsilon] = 0. \quad (2.18)$$

To derive Theorem 1.1 from (2.18), we denote by  $S$  the event that  $\mathcal{G}(n, m)$  consists of  $m$  distinct edges, or, equivalently, that no cycles of length 2 exist in  $\mathcal{G}(n, m)$ . Given that  $S$  occurs,  $\mathcal{G}(n, m)$  is identical to  $G(n, m)$  and  $W'$  is identical to  $W$ . Furthermore, Fact 1.3 implies that  $\mathbb{P}[S] = \Omega(1)$ . Consequently, (2.18) yields

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \mathbb{P}[|\ln Z_k(\mathcal{G}(n, m)) - \ln \mathbb{E}[Z_k(\mathcal{G}(n, m))]| - W'| > \varepsilon | S] \\ &= \lim_{n \rightarrow \infty} \mathbb{P}[|\ln Z_k(G(n, m)) - \ln \mathbb{E}[Z_k(G(n, m))]| - W| > \varepsilon]. \end{aligned} \quad (2.19)$$

As Lemma 3.1 implies that  $\mathbb{E}[Z_k(\mathcal{G}(n, m))], \mathbb{E}[Z_k(G(n, m))] = \Theta(k^n (1 - 1/k)^m)$ , we have  $\mathbb{E}[Z_k(\mathcal{G}(n, m))] = \Theta(\mathbb{E}[Z_k(G(n, m))])$  and with (2.19) it follows that

$$\lim_{n \rightarrow \infty} \mathbb{P}[|\ln Z_k(G(n, m)) - \ln \mathbb{E}[Z_k(G(n, m))]| - W| > \varepsilon] = 0,$$

which proves Theorem 1.1.  $\square$

### 3. THE FIRST MOMENT

The aim in this section is to prove Proposition 2.1. The calculations that have to be done follow the path beaten in [3, 13, 17, 26] and are in fact very similar to [6]. Thus, most of the proofs are deferred to the appendix. Furthermore, at the end of the section we state a result that we need for Proposition 2.4.

Let  $Z_{k,\rho}(G)$  be the number of  $k$ -colourings of the graph  $G$  with colour density  $\rho$ . Let  $\rho^*$  be a  $k$ -dimensional vector with all entries set to  $1/k$ . We define

$$f_1 : \rho \in \mathcal{A}_k \mapsto \mathcal{H}(\rho) + \frac{d}{2} \ln \left( 1 - \sum_{i=1}^k \rho_i^2 \right).$$



In order to determine the expectation of  $Z_{k,\rho}$ , we have to analyse the function  $f_1(\rho)$ . The following lemma was already obtained in [6] and its proof can be found in Section A.1.

**Lemma 3.1.** *Let  $k \geq 3$  and  $d' \in (0, \infty)$ . Then there exist numbers  $C_1 = C_1(k, d)$ ,  $C_2 = C_2(k, d) > 0$  such that for any  $\rho \in \mathcal{A}_k(n)$  we have*

$$C_1 n^{\frac{1-k}{2}} \exp[nf_1(\rho)] \leq \mathbb{E}[Z_{k,\rho}(\mathcal{G}(n, m))] \leq C_2 \exp[nf_1(\rho)]. \quad (3.1)$$

Moreover, if  $\|\rho - \rho^*\|_2 = o(1)$  and  $d = 2m/n$ , then

$$\mathbb{E}[Z_{k,\rho}(\mathcal{G}(n, m))] \sim (2\pi n)^{\frac{1-k}{2}} k^{k/2} \exp[d/2 + nf_1(\rho)]. \quad (3.2)$$

We can now state the expectation of  $Z_k$ . The proof will be carried out in detail in Section A.1.

**Corollary 3.2.** *For any  $k \geq 3$ ,  $d' \in (0, \infty)$  and  $d = 2m/n$ , we have*

$$\mathbb{E}[Z_k(\mathcal{G}(n, m))] \sim \exp[d/2 + nf_1(\rho^*)] \left(1 + \frac{d}{k-1}\right)^{-\frac{k-1}{2}}.$$

*Proof of Proposition 2.1.* The first assertion is immediate from Corollary 3.2. Moreover, the second assertion follows from Corollary 3.2 and the second part of Lemma 3.1.  $\square$

Finally, as our approach requires the analysis of the random variables  $Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))$ , we derive an expression for  $\mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))]$  that we will need to prove Proposition 2.4.

**Lemma 3.3.** *Let  $k \geq 3$ ,  $\omega, \nu \in N$ ,  $d' \in (0, \infty)$  and  $d = 2m/n$ . For  $s \in S_{k,\omega,\nu}$  and  $\rho^{k,\omega,\nu,s}$  as defined in (2.2), we have*

$$\mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))] \sim_\nu |\mathcal{A}_{k,\omega,\nu}^s(n)| (2\pi n)^{\frac{1-k}{2}} k^{k/2} \exp[d/2 + nf_1(\rho^{k,\omega,\nu,s})].$$

*Proof.* Using a Taylor expansion of  $f_1(\rho)$  around  $\rho = \rho^{k,\omega,\nu,s}$ , we get

$$f_1(\rho) = f_1(\rho^{k,\omega,\nu,s}) + \Theta\left(\frac{\omega}{\sqrt{n}}\right) \|\rho - \rho^{k,\omega,\nu,s}\|_1 + \Theta(\|\rho - \rho^{k,\omega,\nu,s}\|_2^2). \quad (3.3)$$

As  $\|\rho - \rho^{k,\omega,\nu,s}\|_1 = O\left(\frac{1}{\nu\sqrt{n}}\right)$  for  $\rho \in \mathcal{A}_{k,\omega,\nu}^s(n)$  and  $\|\rho - \rho^{k,\omega,\nu,s}\|_2^2 = O\left(\frac{1}{\nu^2 n}\right)$ , we conclude that  $f_1(\rho) = f_1(\rho^{k,\omega,\nu,s}) + O\left(\frac{\omega}{\nu n}\right)$  and as this is independent of  $\rho$ , the assertion follows by inserting (3.3) in (3.2) and multiplying by  $|\mathcal{A}_{k,\omega,\nu}^s(n)|$ .  $\square$

#### 4. THE SECOND MOMENT

The aim of this section is to prove Proposition 2.4, which constitutes the main technical contribution of this work and Proposition 2.5, which is done in the last subsection and is based on and an enhancement of results derived in [3]. The crucial points in our analysis are that, similar to [6, 26], we need an asymptotically tight expression for the second moment and instead of confining ourselves to the case of colourings whose colour densities are  $(O(1), n)$ -balanced, as done in most of prior work [3, 7, 13, 17], we need to deal with  $(\omega, n)$ -balanced colour densities for a diverging function  $\omega = \omega(n) \rightarrow \infty$ . However, our work has to extend the calculations from [6] following the example of [26], because we aim for a statement about the whole distribution of  $\ln Z_k(G(n, m))$ . Our line of argument follows that of [26], where analogue statements are proven for the problem of hypergraph 2-colouring.

**4.1. Classifying the overlap.** To standardise the notation, we define the *overlap matrix*  $\rho(\sigma, \tau) = (\rho_{ij}(\sigma, \tau))_{i,j \in [k]}$  for two colour assignments  $\sigma, \tau : [n] \rightarrow [k]$  as the doubly stochastic  $k \times k$ -matrix with entries

$$\rho_{ij}(\sigma, \tau) = \frac{1}{n} \cdot |\sigma^{-1}(i) \cap \tau^{-1}(j)|.$$

We let  $\mathcal{B}_k(n)$  denote the set of all overlap matrices and  $\mathcal{B}_k$  denote the set of all probability measures  $\rho = (\rho_{ij})_{i,j \in [k]}$  on  $[k] \times [k]$ . Moreover, we let  $\bar{\rho}$  signify the  $k \times k$ -matrix with all entries equal to  $k^{-2}$ , the barycentre of  $\mathcal{B}_k$ . For a

$k \times k$ -matrix  $\rho = (\rho_{ij})$ , we introduce the shorthands

$$\rho_{i\star} = \sum_{j=1}^k \rho_{ij}, \quad \rho \cdot \star = (\rho_{i\star})_{i \in [k]}, \quad \rho_{\star j} = \sum_{i=1}^k \rho_{ij}, \quad \rho_{\star \cdot} = (\rho_{\star i})_{i \in [k]}.$$

With the notation from Section 2, we observe that for any  $\sigma, \tau : [n] \rightarrow [k]$  we have  $\rho \cdot \star, \rho_{\star \cdot} \in \mathcal{A}_k(n)$ . We introduce the set

$$\mathcal{B}_{k,\omega}(n) = \left\{ \rho \in \mathcal{B}_k(n) : \rho_{i\star}, \rho_{\star i} \in \left[ \frac{1}{k} - \frac{w}{\sqrt{n}}, \frac{1}{k} + \frac{w}{\sqrt{n}} \right] \text{ for all } i \in [k] \right\},$$

which corresponds to  $\mathcal{A}_{k,\omega}(n)$  insofar as for  $\rho \in \mathcal{B}_{k,\omega}(n)$  we have  $\rho_{i\star}, \rho_{\star i} \in \mathcal{A}_{k,\omega}(n)$  for all  $i \in [k]$ . We remember  $S_{k,\omega,\nu}$  from (2.1). Then for  $s \in S_{k,\omega,\nu}$  we define

$$\mathcal{B}_{k,\omega,\nu}^s(n) = \left\{ \rho \in \mathcal{B}_{k,\omega}(n) : \rho_{i\star}, \rho_{\star i} \in \left[ \rho_i^{k,\omega,\nu,s} - \frac{1}{\nu\sqrt{n}}, \rho_i^{k,\omega,\nu,s} + \frac{1}{\nu\sqrt{n}} \right] \text{ for all } i \in [k] \right\}.$$

Thus, for any fixed  $\nu$ ,  $\mathcal{B}_{k,\omega}(n)$  is a disjoint union of all  $\mathcal{B}_{k,\omega,\nu}^s(n)$  for  $s \in S_{k,\omega,\nu}$ . For a given graph  $G$  on  $[n]$ , we let  $Z_{k,\rho}^{(2)}(G)$  be the number of pairs  $(\sigma, \tau)$  of  $k$ -colourings of  $G$  whose overlap is  $\rho$ . By the linearity of expectation,

$$\mathbb{E} [Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))^2] = \sum_{\rho \in \mathcal{B}_{k,\omega,\nu}^s(n)} \mathbb{E} [Z_{k,\rho}^{(2)}(\mathcal{G}(n, m))]^2. \quad (4.1)$$

To proceed calculating this quantity, we first need the following elementary estimates whose proofs can be found in Section A.2.

**Fact 4.1.** *For any  $k \geq 3$ ,  $d' \in (0, \infty)$  and  $d = 2m/n$ , the following estimates are true.*

(1) *Let  $\rho \in \mathcal{B}_k(n)$ . Then*

$$\mathbb{E} [Z_{k,\rho}^{(2)}(\mathcal{G}(n, m))] \sim \frac{\sqrt{2\pi n}^{\frac{1-k^2}{2}}}{\prod_{i,j=1}^k \sqrt{2\pi \rho_{ij}}} \exp \left[ d/2 + n\mathcal{H}(\rho) + m \ln(1 - \|\rho \cdot \star\|_2^2 - \|\rho_{\star \cdot}\|_2^2 + \|\rho\|_2^2) \right]. \quad (4.2)$$

(2) *For any  $\rho \in \mathcal{B}_k(n)$  with  $\|\rho - \bar{\rho}\|_2^2 = o(1)$ , we have*

$$\mathbb{E} [Z_{k,\rho}^{(2)}(\mathcal{G}(n, m))] \sim k^{k^2} (2\pi n)^{\frac{1-k^2}{2}} \exp \left[ d/2 + n\mathcal{H}(\rho) + m \ln(1 - \|\rho \cdot \star\|_2^2 - \|\rho_{\star \cdot}\|_2^2 + \|\rho\|_2^2) \right]. \quad (4.3)$$

To simplify the notation, we introduce the function  $f_2 : \mathcal{B}_k \rightarrow \mathbb{R}$  defined as

$$f_2(\rho) = \mathcal{H}(\rho) + \frac{d}{2} \ln(1 - \|\rho \cdot \star\|_2^2 - \|\rho_{\star \cdot}\|_2^2 + \|\rho\|_2^2). \quad (4.4)$$

A direct consequence of Fact 4.1 that will be used in the sequel is that for every  $\rho \in \mathcal{B}_k(n)$  we have

$$\mathbb{E} [Z_{k,\rho}^{(2)}(\mathcal{G}(n, m))] = \exp [nf_2(\rho) + O(\ln n)]. \quad (4.5)$$

**4.2. Dividing up the hypercube.** To proceed, we refine equation (4.1). For each  $\omega, \nu \in N$ ,  $s \in S_{k,\omega,\nu}$  and  $\eta > 0$ , we introduce

$$\mathcal{B}_{k,\omega,\nu,\eta}^s(n) = \{ \rho \in \mathcal{B}_{k,\omega,\nu}^s(n) : \|\rho - \bar{\rho}\|_2 \leq \eta \}.$$

We are going to show that the r.h.s. of (4.1) is dominated by the contributions with  $\rho$  “close to”  $\bar{\rho}$  in terms of the euclidean norm. More precisely, for a graph  $G$  let

$$Z_{k,\omega,\nu,\eta}^s(G) = \sum_{\rho \in \mathcal{B}_{k,\omega,\nu,\eta}^s(n)} Z_{k,\rho}^{(2)}(G) \quad \text{for any } \eta > 0.$$

Then the second moment argument performed in [3] fairly directly yields the following statement showing that overlap matrices that are far apart from  $\bar{\rho}$  do asymptotically not contribute to the second moment.

**Proposition 4.2.** *Assume that  $k \geq 3$  and  $d' < 2(k-1) \ln(k-1)$ . Further, let  $\omega, \nu \in N$ . Then for any fixed  $\eta > 0$  and any  $s \in S_{k,\omega,\nu}$ , it holds that*

$$\mathbb{E} [Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))^2] \sim \mathbb{E} [Z_{k,\omega,\nu,\eta}^s(\mathcal{G}(n, m))]^2.$$

To prove this proposition, we first define a function

$$\bar{f}_2 : \rho \in \mathcal{B}_{k,\omega}(n) \rightarrow \mathbb{R}, \quad \rho \mapsto \mathcal{H}(\rho) + \frac{d}{2} \ln \left( 1 - \frac{2}{k} + \|\rho\|_2^2 \right).$$

The following lemma shows how  $f_2$  defined in (4.4) relates to  $\bar{f}_2$ .

**Lemma 4.3.** *For  $\rho = (\rho_{ij}) \in \mathcal{B}_{k,\omega}(n)$ , we have*

$$\exp [nf_2(\rho)] \sim \exp [n\bar{f}_2(\rho) + O(\omega^2)].$$

*Proof.* We define the function

$$\zeta(\rho) = f_2(\rho) - \bar{f}_2(\rho)$$

and derive an upper bound on  $\zeta(\rho)$ . By definition, for each  $\rho \in \mathcal{B}_{k,\omega}(n)$  there exist  $\alpha = (\alpha_i)_{i \in [k]}$  and  $\beta = (\beta_j)_{j \in [k]}$  such that  $\rho_{i*} = \frac{1}{k} + \alpha_i$  and  $\rho_{*j} = \frac{1}{k} + \beta_j$  for all  $i, j \in [k]$  with  $|\alpha_i|, |\beta_j| \leq \frac{\omega}{\sqrt{n}}$ . Thus,

$$f_2(\rho) = \mathcal{H}(\rho) + \frac{d}{2} \ln (1 - \|\bar{\rho} \cdot \star + \alpha\|_2^2 - \|\bar{\rho} \star \cdot + \beta\|_2^2 + \|\rho\|_2^2).$$

As we are only interested in the difference between  $f_2$  and  $\bar{f}_2$ , we can reparametrise  $\zeta$  as

$$\zeta(\alpha, \beta) = \frac{d}{2} \ln \left( \frac{1 - \|\bar{\rho} \cdot \star + \alpha\|_2^2 - \|\bar{\rho} \star \cdot + \beta\|_2^2 + \|\rho\|_2^2}{1 - \frac{2}{k} + \|\rho\|_2^2} \right).$$

Differentiating and simplifying the expression yields  $\frac{\partial \zeta}{\partial \alpha_i}(\alpha, \beta), \frac{\partial \zeta}{\partial \beta_j}(\alpha, \beta) = O\left(\frac{\omega}{\sqrt{n}}\right)$  for all  $i, j \in [k]$ . According to the fundamental theorem of calculus, it follows that

$$\max_{\rho \in \mathcal{B}_{k,\omega}(n)} |\zeta(\rho)| = \int_{-\omega/\sqrt{n}}^{\omega/\sqrt{n}} O\left(\frac{\omega}{\sqrt{n}}\right) d\alpha_1 = O\left(\frac{\omega^2}{n}\right),$$

completing the proof.  $\square$

*Proof of Proposition 4.2.* Equation (4.5) combined with Lemma 4.3 reduces our task to studying the function  $\bar{f}_2(\rho)$ . For the range of  $d$  covered by Proposition 4.2, this analysis is the main technical achievement of [3], where (essentially) the following statement is proved.

**Lemma 4.4.** *Assume that  $k \geq 3, \omega \in N$  as well as  $d' \leq 2(k-1) \ln(k-1)$  and  $d = 2m/n$ . For any  $n > 0$  and any overlap matrix  $\rho \in \mathcal{B}_{k,\omega}(n)$ , we have*

$$\bar{f}_2(\rho) \leq \bar{f}_2(\bar{\rho}) - \frac{2(k-1) \ln(k-1) - d}{4(k-1)^2} (k^2 \|\rho\|_2^2 - 1) + o(1). \quad (4.6)$$

*Proof.* For  $\rho$  such that  $\sum_{i=1}^k \rho_{ij} = \sum_{j=1}^k \rho_{ji} = 1/k$ , the bound (4.6) is proved in [3, Section 3]. This implies that (4.6) also holds for  $\rho \in \mathcal{B}_{k,\omega}(n)$ , because  $\bar{f}_2$  is uniformly continuous on the compact set  $\mathcal{B}_{k,\omega}(n)$ .  $\square$

Now, assume that  $k$  and  $d$  satisfy the assumptions of Proposition 4.2 and let  $\nu \in N$  and  $\eta > 0$  be any fixed number. Then, for any  $\hat{\rho} \in \mathcal{B}_{k,\omega,\nu}^s(n)$ , we have  $\|\hat{\rho} - \bar{\rho}\|_2 = O\left(\frac{\omega}{\sqrt{n}}\right)$ . Consequently, we obtain with (4.5) that

$$\sum_{\substack{\rho \in \mathcal{B}_{k,\omega,\nu}^s(n) \\ \|\rho - \bar{\rho}\|_2 \leq \eta}} \mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] \geq \mathbb{E} \left[ Z_{k,\hat{\rho}}^{(2)}(\mathcal{G}(n, m)) \right] \geq \exp [nf_2(\bar{\rho}) + O(\ln n)]. \quad (4.7)$$

On the other hand, the function  $\mathcal{B} \rightarrow \mathbb{R}, \rho \mapsto k^2 \|\rho\|_2^2$  is smooth, strictly convex and attains its global minimum of 1 at  $\rho = \bar{\rho}$ . Consequently, there exist  $(c_k)_k > 0$  such that if  $\|\rho - \bar{\rho}\|_2 > \eta$ , then  $(k^2 \|\rho\|_2^2 - 1) \geq c_k$ . Hence, Fact 4.1, Lemma 4.3 and Lemma 4.4 yield

$$\sum_{\substack{\rho \in \mathcal{B}_{k,\omega,\nu}^s(n) \\ \|\rho - \bar{\rho}\|_2 > \eta}} \mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] \leq \exp [nf_2(\bar{\rho}) - nc_k d_k + o(n)], \quad \text{where } d_k = \frac{2(k-1) \ln(k-1) - d}{4(k-1)^2} > 0. \quad (4.8)$$

Combining (4.8) and (4.7), we conclude that  $\mathbb{E} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))^2 \right] \sim \mathbb{E} \left[ Z_{k,\omega,\nu,\eta}^{s(2)}(\mathcal{G}(n, m)) \right]$ , thereby completing the proof of Proposition 4.2.  $\square$

Having reduced our task to studying overlaps  $\rho$  such that  $\|\rho - \bar{\rho}\|_2 \leq \eta$  for a small but fixed  $\eta > 0$ , in this section we are going to argue that, in fact, it suffices to consider  $\rho$  such that  $\|\rho - \bar{\rho}\|_2 \leq n^{-3/8}$  (where the constant  $3/8$  is somewhat arbitrary; any number smaller than  $1/2$  would do). More precisely, we have

**Proposition 4.5.** *Assume that  $k \geq 3$  and that  $d' < d_{k,\text{cond}}$ . Let  $\nu, \omega \in N$  and  $s \in S_{k,\omega,\nu}$ . There exists a number  $\eta_0 = \eta_0(d', k)$  such that for any  $0 < \eta < \eta_0$  we have*

$$\mathbb{E} \left[ Z_{k,\omega,\nu,\eta}^{s(2)}(\mathcal{G}(n, m)) \right] \sim \mathbb{E} \left[ Z_{k,\omega,\nu,n^{-3/8}}^{s(2)}(\mathcal{G}(n, m)) \right].$$

The key to proving this proposition is the following lemma. It specifies the expected number of pairs of solutions in the cases where the overlap matrices  $\rho \in \mathcal{B}_{k,\omega,\nu}^s(n)$  satisfy  $\|\rho - \bar{\rho}\|_2 \leq n^{-3/8}$  or  $\|\rho - \bar{\rho}\|_2 \in (n^{-3/8}, \eta)$ .

**Lemma 4.6.** *Let  $k \geq 3$ ,  $d' < (k-1)^2$  and  $d = 2m/n$ . Set*

$$C_n(d, k) = \exp[d/2] k^{k^2} (2\pi n)^{\frac{1-k^2}{2}} \quad \text{and} \quad D(d, k) = k^2 \left( 1 - \frac{d}{(k-1)^2} \right). \quad (4.9)$$

- If  $\rho \in \mathcal{B}_{k,\omega,\nu,\eta}^s(n)$  satisfies  $\|\rho - \bar{\rho}\|_2 \leq n^{-3/8}$ , then

$$\mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] \sim C_n(d, k) \exp \left[ 2nf_1(\rho^*) - n \frac{D(d, k)}{2} \|\rho - \bar{\rho}\|_2^2 \right]. \quad (4.10)$$

- There exist numbers  $\eta = \eta(d, k) > 0$  and  $A = A(d, k) > 0$  such that if  $\rho \in \mathcal{B}_{k,\omega,\nu,\eta}^s(n)$  satisfies  $\|\rho - \bar{\rho}\|_2 \in (n^{-3/8}, \eta)$ , then

$$\mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] = \exp \left[ 2nf_1(\rho^*) - An^{1/4} \right]. \quad (4.11)$$

*Proof.* As Fact 4.1 yields  $\mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] \sim C_n(d, k) \exp[nf_2(\rho)]$ , we have to analyse  $f_2$ . Expanding this function around  $\bar{\rho}$  yields

$$f_2(\rho) = f_2(\bar{\rho}) - \frac{D(d, k)}{2} \|\rho - \bar{\rho}\|_2^2 + O(\|\rho - \bar{\rho}\|_2^3). \quad (4.12)$$

Consequently, for  $\|\rho - \bar{\rho}\|_2 \leq n^{-3/8}$ ,

$$\exp[nf_2(\rho)] = \exp \left[ nf_2(\bar{\rho}) - n \frac{D(d, k)}{2} \|\rho - \bar{\rho}\|_2^2 + O(n^{-1/8}) \right].$$

As  $f_2$  satisfies  $f_2(\bar{\rho}) = 2f_1(\rho^*)$ , the statement in (4.10) follows.

To prove (4.11), we observe that similarly to (4.12) and because  $f_2$  is smooth in a neighbourhood of  $\bar{\rho}$ , there exist  $\eta > 0$  and  $A > 0$  such that for  $\|\rho - \bar{\rho}\|_2 \leq \eta$ ,

$$f_2(\rho) \leq f_2(\bar{\rho}) - A\|\rho - \bar{\rho}\|_2^2.$$

Hence, if  $\|\rho - \bar{\rho}\|_2 \in (n^{-3/8}, \eta)$ , then

$$\mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] = O \left( n^{\frac{1-k^2}{2}} \right) \exp[nf_2(\rho)] \leq \exp \left[ 2nf_1(\rho^*) - An^{1/4} \right],$$

as claimed.  $\square$

*Proof of Proposition 4.5.* We fix  $s \in S_{k,\omega,\nu}$ . Further, we fix  $\eta > 0$  and  $A > 0$  as given by Lemma 4.6. For each  $\hat{\rho} \in \mathcal{B}_{k,\omega,\nu,\eta}^s(n)$ , we have  $\|\hat{\rho} - \bar{\rho}\|_2 = O\left(\frac{\omega}{\sqrt{n}}\right)$  and obtain from the first part of Lemma 4.6 that

$$\mathbb{E} \left[ Z_{k,\omega,\nu,n^{-3/8}}^{s(2)}(\mathcal{G}(n, m)) \right] \geq \mathbb{E} \left[ Z_{k,\rho_0}^{(2)}(\mathcal{G}(n, m)) \right] \sim C_n(d, k) \exp \left[ 2nf_1(\rho^*) + O(\omega^2) \right]. \quad (4.13)$$

On the other hand, because  $|\mathcal{B}_{k,\omega,\nu,\eta}^s(n)|$  is bounded by a polynomial in  $n$ , the second part of Lemma 4.6 yields

$$\sum_{\substack{\rho \in \mathcal{B}_{k,\omega,\nu,\eta}^s(n) \\ \|\rho - \bar{\rho}\|_2 > n^{-3/8}}} \mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] \leq \exp \left[ 2nf_1(\rho^*) - An^{1/6} + O(\ln n) \right]. \quad (4.14)$$

Combining (4.13) and (4.14), we obtain

$$\mathbb{E} \left[ Z_{k,\omega,\nu,\eta}^{s(2)}(\mathcal{G}(n, m)) \right] \sim \sum_{\rho \in \mathcal{B}_{k,\omega,\nu,n-3/8}^s(n)} \mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] \sim \mathbb{E} \left[ Z_{k,\omega,\nu,n-3/8}^{s(2)}(\mathcal{G}(n, m)) \right],$$

as claimed.  $\square$

**4.3. Calculating the constant.** This section is dedicated to computing the contribution of the overlap matrices  $\rho \in \mathcal{B}_{k,\omega,\nu,n-3/8}^s(n)$ . To this aim, we first show that in each region of the hypercube we can approximate  $f_2$  by a function where the marginals are set to those of the centre of this region as defined in (2.2). More formally, let  $f_2^s : \mathcal{B}_k \rightarrow \mathbb{R}$  be defined as

$$f_2^s : \rho \mapsto \mathcal{H}(\rho) + \frac{d}{2} \ln (1 - 2\|\rho^{k,\omega,\nu,s}\|_2^2 + \|\rho\|_2^2).$$

Then the following is true

**Lemma 4.7.** *Let  $k \geq 3, \omega, \nu \in N$  and  $C_n(d, k)$  as in (4.9). Then for  $\rho \in \mathcal{B}_{k,\omega,\nu,n-3/8}^s(n)$  it holds that*

$$\mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] \sim C_n(d, k) \exp \left[ n f_2^s(\rho) + O\left(\frac{\omega}{\nu}\right) \right].$$

*Proof.* Equation (4.3) of Fact 4.1 yields that

$$\mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] \sim C_n(d, k) \exp [n f_2(\rho)]. \quad (4.15)$$

For  $s \in S_{k,\omega,\nu}$ , we define the function

$$\zeta^s(\rho) = f_2(\rho) - f_2^s(\rho).$$

To derive an upper bound on  $\zeta^s(\rho)$  for all values  $\rho \in \mathcal{B}_{k,\omega,\nu,n-3/8}^s(n)$ , we first we observe that there exist  $\alpha = (\alpha_i)_{i \in [k]}$  and  $\beta = (\beta_j)_{j \in [k]}$  such that the function  $f_2$  can be expressed by setting  $\rho_{i*} = \rho_i^{k,\omega,\nu,s} + \alpha_i$  and  $\rho_{*j} = \rho_j^{k,\omega,\nu,s} + \beta_j$  for all  $i, j \in [k]$  with  $|\alpha_i|, |\beta_j| \leq \frac{1}{\nu\sqrt{n}}$ . Thus,

$$f_2 : \rho \mapsto \mathcal{H}(\rho) + \frac{d}{2} \ln (1 - \|\rho^{k,\omega,\nu,s} + \alpha\|_2^2 - \|\rho^{k,\omega,\nu,s} + \beta\|_2^2 + \|\rho\|_2^2).$$

As we are only interested in the difference between  $f_2$  and  $f_2^s$ , we can reparametrise  $\zeta^s$  as

$$\zeta^s(\alpha, \beta) = \frac{d}{2} \ln \left( \frac{1 - \|\rho^{k,\omega,\nu,s} + \alpha\|_2^2 - \|\rho^{k,\omega,\nu,s} + \beta\|_2^2 + \|\rho\|_2^2}{1 - 2\|\rho^{k,\omega,\nu,s}\|_2^2 + \|\rho\|_2^2} \right).$$

Differentiating and simplifying the expression yields  $\frac{\partial \zeta^s}{\partial \alpha_i}(\alpha, \beta), \frac{\partial \zeta^s}{\partial \beta_j}(\alpha, \beta) = O\left(\frac{\omega}{\sqrt{n}}\right)$  for all  $i, j \in [k]$ . According to the fundamental theorem of calculus it follows for every  $s \in S_{k,\omega,\nu}$  that

$$\max_{\rho \in \mathcal{B}_{k,\omega,\nu,n-3/8}^s(n)} |\zeta^s(\rho)| = \int_{-(\nu\sqrt{n})^{-1}}^{(\nu\sqrt{n})^{-1}} O\left(\frac{\omega}{\sqrt{n}}\right) d\alpha_1 = O\left(\frac{\omega}{n\nu}\right).$$

Combining this with (4.15) yields the assertion.  $\square$

Now we are able to give a very precise expression for the second moment.

**Proposition 4.8.** *Assume that  $k \geq 3, \omega, \nu \in N, d' < (k-1)^2$  and  $d = 2m/n$ . Let  $s \in S_{k,\omega,\nu}$ . Then*

$$\begin{aligned} & \mathbb{E} \left[ Z_{k,\omega,\nu,n-3/8}^{s(2)}(\mathcal{G}(n, m)) \right] \\ & \sim_{\nu} \left( |\mathcal{A}_{k,\omega}(n)| (2\pi n)^{\frac{1-k}{2}} k^{k/2} \exp [n f_1(\rho^{k,\omega,\nu,s})] \right)^2 \exp [d/2] \left( 1 - \frac{d}{(k-1)^2} \right)^{-\frac{(k-1)^2}{2}}. \end{aligned}$$

The rest of this subsection will be dedicated to proving this proposition. In due course we are going to need the set of matrices with coefficients in  $\frac{1}{n}\mathbb{Z}$  whose lines and columns sum to zero:

$$\mathcal{E}_n = \left\{ (\epsilon_{i,j})_{\substack{1 \leq i \leq k \\ 1 \leq j \leq k}}, \forall i, j \in [k], \epsilon_{i,j} \in \frac{1}{n}\mathbb{Z}, \forall j \in [k], \sum_{i=1}^k \epsilon_{ij} = \sum_{i=1}^k \epsilon_{ji} = 0 \right\}. \quad (4.16)$$

The following result regards Gaussian summations over matrices in  $\mathcal{E}_n$ .

**Lemma 4.9.** *Let  $k \geq 2$ ,  $d' < (k-1)^2$  and  $D > 0$  be fixed. Then*

$$\sum_{\epsilon \in \mathcal{E}_n} \exp \left[ -n \frac{D}{2} \|\epsilon\|_2^2 + o(n^{1/2}) \|\epsilon\|_2 \right] \sim \left( \sqrt{2\pi n} \right)^{(k-1)^2} D^{-\frac{(k-1)^2}{2}} k^{-(k-1)}.$$

Lemma 4.9 and its proof are very similar to an argument used in [17, Section 3]. In fact, Lemma 4.9 follows from

**Lemma 4.10** ([17, Lemma 6 (b) and 7 (c)]). *There is a  $(k-1)^2 \times (k-1)^2$ -matrix  $\mathcal{H} = (\mathcal{H}_{(i,j),(i',j')})_{i,j,i',j' \in [k-1]}$  such that for any  $\varepsilon = (\varepsilon_{ij})_{i,j \in [k]} \in \mathcal{E}_n$  we have*

$$\sum_{i,j,i',j' \in [k-1]} \mathcal{H}_{(i,j),(i',j')} \varepsilon_{ij} \varepsilon_{i'j'} = \|\varepsilon\|_2^2.$$

*This matrix  $\mathcal{H}$  is positive definite and  $\det \mathcal{H} = k^{2(k-1)}$ .*

The Proof of Lemma 4.9 can be found in Section A.2.

Now we are ready to prove Proposition 4.8.

*Proof of Proposition 4.8.* Lemma 4.7 states that for every  $\rho \in \mathcal{B}_{k,\omega,\nu,n-3/8}^s(n)$  we have

$$\mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] \sim C_n(d, k) \exp \left[ n f_2^s(\rho) + O \left( \frac{\omega}{\nu} \right) \right]. \quad (4.17)$$

Thus, all we have to do is analysing the function  $f_2^s$  for  $s \in S_{k,\omega,\nu}$ . To this aim, we expand  $f_2^s(\rho)$  around  $\rho = \rho^s$  where  $\rho^s = (\rho_{ij}^s)_{i,j}$  with  $\rho_{ij}^s = \rho_i^{k,\omega,\nu,s} \cdot \rho_j^{k,\omega,\nu,s}$ . Then with  $D(d, k)$  as defined in (4.9) we have

$$f_2^s(\rho) = f_2^s(\rho^s) + \Theta \left( \frac{\omega}{n} \right) \|\rho - \rho^s\|_2 - \frac{D(d, k)}{2} \|\rho - \rho^s\|_2^2 + o(n^{-1}). \quad (4.18)$$

Combining (4.18) with (4.17), we find that

$$\mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] \sim C_n(d, k) \exp \left[ n f_2^s(\rho^s) + \Theta(\omega) \|\rho - \rho^s\|_2 - n \frac{D(d, k)}{2} \|\rho - \rho^s\|_2^2 + O \left( \frac{\omega}{\nu} \right) \right]. \quad (4.19)$$

For two vectors of “marginals”  $\rho^0, \rho^1 \in \mathcal{B}_{k,\omega,\nu}^s(n)$ , we introduce the set of overlap matrices

$$\mathcal{B}_{k,\omega,\nu,n-3/8}^s(n, \rho^0, \rho^1) = \{\rho \in \mathcal{B}_{k,\omega,\nu,n-3/8}^s(n) : \rho \cdot \star = \rho^0, \rho_{\star \cdot} = \rho^1\}.$$

and observe that with this definition we have

$$\mathbb{E} \left[ Z_{k,\omega,\nu,n-3/8}^{s(2)}(\mathcal{G}(n, m)) \right] = \sum_{\rho^0, \rho^1 \in \mathcal{B}_{k,\omega,\nu}^s(n)} \sum_{\rho \in \mathcal{B}_{k,\omega,\nu,n-3/8}^s(n, \rho^0, \rho^1)} \mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right]. \quad (4.20)$$

In particular, the set  $\mathcal{B}_{k,\omega,\nu,n-3/8}^s(n, \rho^0, \rho^1)$  contains the “product” overlap  $\rho^0 \otimes \rho^1$  defined by  $(\rho^0 \otimes \rho^1)_{ij} = \rho_i^0 \rho_j^1$  for  $i, j \in [k]$ . To proceed, we fix two colour densities  $\rho^0, \rho^1 \in \mathcal{B}_{k,\omega,\nu}^s(n)$  and simplify the notation by writing

$$\widehat{\mathcal{B}} = \mathcal{B}_{k,\omega,\nu,n-3/8}^s(n, \rho^0, \rho^1), \quad \widehat{\rho} = \rho^0 \otimes \rho^1.$$

Thus, the inner sum from (4.20) simplifies to

$$\mathcal{S}_1 = \sum_{\rho \in \widehat{\mathcal{B}}} \mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right].$$

and we are going to evaluate this quantity. We observe that with  $\mathcal{E}_n$  as defined in (4.16), for each  $\rho \in \widehat{\mathcal{B}}$  we can find  $\varepsilon \in \mathcal{E}_n$  such that

$$\rho = \widehat{\rho} + \varepsilon.$$

Hence, this gives  $\|\rho - \rho^s\|_2 = \|\widehat{\rho} + \varepsilon - \rho^s\|_2$  and the triangle inequality yields

$$\|\varepsilon\|_2 - \|\widehat{\rho} - \rho^s\|_2 \leq \|\widehat{\rho} + \varepsilon - \rho^s\|_2 \leq \|\varepsilon\|_2 + \|\widehat{\rho} - \rho^s\|_2.$$

By definition of  $\widehat{\rho}$  and  $\rho^s$ , we have  $\|\widehat{\rho} - \rho^s\|_2 \leq \frac{1}{\nu\sqrt{n}}$  and consequently

$$\|\rho - \rho^s\|_2 = \|\varepsilon\|_2 + O \left( \frac{1}{\nu\sqrt{n}} \right). \quad (4.21)$$



Observing that  $f_2^s(\rho^s) = (f_1(\rho^{k,\omega,\nu,s}))^2$  and inserting (4.21) into (4.19) while taking first  $n \rightarrow \infty$  and afterwards  $\nu \rightarrow \infty$ , we obtain

$$\mathcal{S}_1 \sim_\nu C_n(d, k) \exp[2nf_1^s(\rho^{k,\omega,\nu,s})] \sum_{\rho \in \widehat{\mathcal{B}}} \exp\left[-n \frac{D(d, k)}{2} \|\varepsilon\|_2^2 + o(n^{1/2}) \|\varepsilon\|_2\right]. \quad (4.22)$$

To apply Lemma 4.9, we have to relate  $\rho \in \widehat{\mathcal{B}}$  to  $\varepsilon \in \mathcal{E}_n$ . From the definitions we obtain

$$\{\widehat{\rho} + \varepsilon : \varepsilon \in \mathcal{E}_n, \|\varepsilon\|_2 \leq n^{-3/8}/2\} \subset \{\rho \in \widehat{\mathcal{B}}\} \subset \{\widehat{\rho} + \varepsilon : \varepsilon \in \mathcal{E}_n\}.$$

We show that the contribution of  $\varepsilon \in \mathcal{E}_n$  with  $\|\varepsilon\|_2 > n^{-3/8}/2$  is negligible:

$$\begin{aligned} \mathcal{S}_2 &= C_n(d, k) \exp[2nf_1^s(\rho^{k,\omega,\nu,s})] \sum_{\substack{\varepsilon \in \mathcal{S}_n \\ \|\varepsilon\|_2 > n^{-3/8}/2}} \exp\left[-n \frac{D(d, k)}{2} \|\varepsilon\|_2^2 (1 + o(1))\right] \\ &= C_n(d, k) \exp[2nf_1^s(\rho^{k,\omega,\nu,s})] \sum_{\substack{l \in \mathbb{Z}/n \\ l > n^{-3/8}/2}} \sum_{\substack{\varepsilon \in \mathcal{S}_n \\ \|\varepsilon\|_2 = l}} \exp\left[-nl^2 \frac{D(d, k)}{2} (1 + o(1))\right] \\ &= C_n(d, k) \exp[2nf_1^s(\rho^{k,\omega,\nu,s})] O\left(n^{k^2}\right) \exp\left[-\frac{D(d, k)}{2} n^{1/4}\right] \end{aligned}$$

Consequently, (4.22) yields  $\Sigma_2 = o(\Sigma_1)$ . Thus, we obtain from Lemma 4.9 that

$$\begin{aligned} \mathcal{S}_1 &\sim_\nu C_n(d, k) \exp[2nf_1^s(\rho^{k,\omega,\nu,s})] \sum_{\rho \in \widehat{\mathcal{B}}} \exp\left[-n \frac{D(d, k)}{2} \|\varepsilon\|_2^2 + o(n^{1/2}) \|\varepsilon\|_2\right] \\ &\sim_\nu C_n(d, k) \exp[2nf_1^s(\rho^{k,\omega,\nu,s})] \left(\sqrt{2\pi n}\right)^{(k-1)^2} k^{-k(k-1)} \left(1 - \frac{d}{(k-1)^2}\right)^{-\frac{(k-1)^2}{2}}. \end{aligned} \quad (4.23)$$

In particular, the last expression is independent of the choice of the vectors  $\rho^0, \rho^1$  that defined  $\widehat{\mathcal{B}}$ . Therefore, substituting (4.23) in the decomposition (4.20) completes the proof of Proposition 4.8.  $\square$

*Proof of Proposition 2.4.* First observe that

$$\exp\left[\sum_{l \geq 2} \lambda_l \delta_l^2\right] = \left(1 - \frac{d}{(k-1)^2}\right)^{-\frac{(k-1)^2}{2}} \exp\left[-\frac{d}{2}\right].$$

Proposition 2.4 is immediately obtained by combining Lemma 3.3 with Propositions 4.2, 4.5 and 4.8.  $\square$

**4.4. Up to the condensation threshold.** In this last subsection we prove Proposition 2.5. In the regime  $2(k-1) \ln(k-1) \leq d' < d_{k,\text{cond}}$  for  $k \geq k_0$  for some big constant  $k_0$ , we consider random variables  $\widetilde{Z}_{k,\omega,\nu}^s$  instead of  $Z_{k,\omega,\nu}^s$ . To prove the proposition we show the following result by adapting our setting in a way that we can apply the second moments argument from [13] and [7].

**Proposition 4.11.** *Let  $\omega, \nu \in N$ . There is a constant  $k_0 > 3$  such that for  $k \geq k_0$  and  $2(k-1) \ln(k-1) \leq d' < d_{k,\text{cond}}$  the following is true. For each  $s \in S_{k,\omega,\nu}$ , there exists an integer-valued random variable  $0 \leq \widetilde{Z}_{k,\omega,\nu}^s \leq Z_{k,\omega,\nu}^s$  that satisfies*

$$\mathbb{E}\left[\widetilde{Z}_{k,\omega,\nu}^s(\mathcal{G}(n, m))\right] \sim \mathbb{E}\left[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))\right]$$

*and such that for any fixed  $\eta > 0$  we have  $\mathbb{E}\left[\widetilde{Z}_{k,\omega,\nu}^s(\mathcal{G}(n, m))^2\right] \leq (1 + o(1)) \mathbb{E}\left[Z_{k,\omega,\nu,\eta}^{s(2)}(\mathcal{G}(n, m))\right]$ .*

In this section we work with the Erdős-Rényi random graph model  $G(n, p)$ , which is a random graph on  $[n]$  vertices where every possible edge is present with probability  $p = d/n$  independently. We further assume from now on that  $k$  divides  $n$ .

The use of results from [13, 7] is complicated by the fact that we are dealing with  $(\omega, n)$ -balanced  $k$ -colourings that allow a larger discrepancy between the colour classes than [13, 7], where balanced colourings are defined such that in

each color class only a deviation of at most  $\sqrt{n}$  from the typical value  $n/k$  is allowed. To circumvent this problem, we introduce the following:

Choose a map  $\sigma : [n] \rightarrow [k]$  uniformly at random and generate a graph  $G(n, p', \sigma)$  on  $[n]$  by connecting any two vertices  $v, w \in [n]$  such that  $\sigma(v) \neq \sigma(w)$  with probability  $p' = dk/(n(k-1))$  independently.

Given  $\sigma$  and  $G(n, p', \sigma)$ , we define

$$\alpha_i = |\sigma^{-1}(i) - n/k| \quad \text{for } i \in [k]$$

and let  $\alpha = \max_{i \in [k]} \alpha_i$ . Thus, by definition  $\alpha \leq \omega\sqrt{n}$ . We set  $n' = n + k\lceil\alpha\rceil$ . Further, we let

$$\beta_i = |\sigma^{-1}(i) - (n + k\lceil\alpha\rceil)/k| \quad \text{for } i \in [k].$$

We then construct a coloured graph  $G'_{n', p', \sigma'}$  from  $G(n, p', \sigma)$  in the following way:

- Add  $k\lceil\alpha\rceil$  vertices to  $G(n, p)$  and denote them by  $n+1, n+2, \dots, n+k\lceil\alpha\rceil$ .
- Define a colouring  $\sigma' : [n'] \rightarrow [k]$  by setting  $\sigma'(i) = \sigma(i)$  for  $i \in [n]$ ,  $\sigma'(i) = 1$  for  $i \in n+1, \dots, n+\beta_1$  and  $\sigma'(i) = j$  for  $j \in \{2, \dots, k\}$  and  $i \in n+\beta_{j-1}+1, \dots, n+\beta_j$ .
- Add each possible edge  $(i, j)$  with  $\sigma'(i) \neq \sigma'(j)$  involving a vertex  $i \in \{n+1, \dots, n+k\lceil\alpha\rceil\}$  with probability  $p' = dk/(n(k-1))$ .

We call a colouring  $\tau : [n] \rightarrow [k]$  of a graph  $G$  on  $[n]$  *perfectly balanced* if  $|\tau^{-1}(i)| = |\tau^{-1}(j)|$  for all  $i, j \in [k]$  and we denote the set of all such perfectly balanced colourings by  $\tilde{\mathcal{B}}_k(n)$ . Then the following holds by construction:

**Fact 4.12.**  $G'_{n', p', \sigma'}$  has the same distribution as  $G(n', p', \tau)$  conditioned on the event that  $\tau : [n'] \rightarrow k$  is perfectly balanced.

Let  $G''_{n, p', \sigma' | [n]}$  denote the graph obtained from  $G'_{n', p', \sigma'}$  by deleting the vertices  $n+1, \dots, n+k\lceil\alpha\rceil$  and the incident edges.

**Fact 4.13.**  $G''_{n, p', \sigma' | [n]}$  has the same distribution as  $G(n, p', \tau)$  conditioned on the event that  $\tau$  is  $(\omega, n)$ -balanced.

To proceed, we adopt the following notation from [13]: Let  $\rho \in \mathcal{B}_k$  be called *s-stable* if it has precisely  $s$  entries bigger than  $0.51/k$ . Further, let  $\tilde{\mathcal{B}}_k$  be the set of all  $\rho \in \mathcal{B}_k$  such that

$$\sum_{j=1}^k \rho_{ij} = \sum_{j=1}^k \rho_{ji} = 1/k \quad \text{for all } i \in [k].$$

Then any  $\rho \in \tilde{\mathcal{B}}_k$  is *s-stable* for some  $s \in \{0, 1, \dots, k\}$ . In addition, let  $\kappa = \ln^{20} k/k$  and let us call  $\rho \in \mathcal{B}_k$  *separable* if  $k\rho_{ij} \notin (0.51, 1 - \kappa)$  for all  $i, j \in [k]$ . A  $k$ -colouring  $\sigma$  of a graph  $G$  on  $[n]$  is called *separable* if for any other  $k$ -colouring  $\tau$  of  $G$  the overlap matrix  $\rho(\sigma, \tau)$  is separable. We have the following result:

**Lemma 4.14.** Let  $s \in S_{k, \omega, \nu}$ . There is  $k_0 > 0$  such that for all  $k > k_0$  and all  $d'$  such that  $2(k-1)\ln(k-1) \leq d' \leq (2k-1)\ln k$  the following is true. Let  $\tilde{Z}_{k, \omega, \nu}^s(\mathcal{G}(n, m))$  denote the number of  $(\omega, n)$ -balanced  $k$ -colourings of  $\mathcal{G}(n, m)$  that fail to be separable. Then  $\mathbb{E}[\tilde{Z}_{k, \omega, \nu}^s(\mathcal{G}(n, m))] = o(\mathbb{E}[Z_{k, \omega, \nu}^s(\mathcal{G}(n, m))])$ .

To prove this lemma, we combine Fact 4.12 with [13, Lemma 3.3]. This yields the following<sup>1</sup>.

**Lemma 4.15** ([13]). There is  $k_0 > 0$  such that for all  $k \geq k_0$  and all  $d'$  with  $2(k-1)\ln(k-1) \leq d' \leq (2k-1)\ln k$  each  $\tau \in \tilde{\mathcal{B}}_k(n')$  is separable in  $G'_{n', p', \tau}$  w.h.p..

*Proof of Lemma 4.14.* Choose a map  $\sigma : [n] \rightarrow [k]$  uniformly at random and generate a graph  $G(n, p', \sigma)$  on  $[n]$  by connecting any two vertices  $v, w \in [n]$  such that  $\sigma(v) \neq \sigma(w)$  with probability  $p'$  independently. Construct  $G'_{n', p', \sigma'}$  from  $G(n, p', \sigma)$  in the way defined above. Then  $\sigma' \in \tilde{\mathcal{B}}_k(n)$ . By Lemma 4.15,  $\sigma'$  is separable in  $G'_{n', p', \sigma'}$  w.h.p.. Thus,  $\sigma$  is separable in  $G''_{n, p', \sigma' | [n]}$  if we define separability using  $\kappa' = \frac{\ln^{21} k}{k}$ . By choosing  $k_0$  large enough and applying Fact 4.13, the assertion follows.  $\square$

For the next ingredient to the proof of Proposition 4.11, we need the following definition. For a graph  $G$  on  $[n]$  and a  $k$ -colouring  $\sigma$  of  $G$ , we let  $\mathcal{C}(G, \sigma)$  be the set of all  $\tau \in \mathcal{B}_k$  that are  $k$ -colourings of  $G$  such that  $\rho(\sigma, \tau)$  is  $k$ -stable.

<sup>1</sup>As a matter of fact, Lemma 3.2 in [13] also holds for densities  $2(k-1)\ln(k-1) \leq d' \leq 2(k-1)\ln k - 2$ , as all steps in the proof are also valid in this regime.

**Lemma 4.16.** *Let  $s \in S_{k,\omega,\nu}$ . There is  $k_0 > 0$  such that for all  $k > k_0$  and all  $d'$  such that  $(2k-1)\ln k - 2 \leq d' \leq d_{k,\text{cond}}$  the following is true. There exists an  $\varepsilon > 0$  such that if  $\tilde{Z}_{k,\omega,\nu}^s(\mathcal{G}(n, m))$  denotes the number of  $(\omega, n)$ -balanced  $k$ -colourings  $\sigma$  of  $\mathcal{G}(n, m)$  satisfying  $|\mathcal{C}(\mathcal{G}(n, m), \sigma)| > \mathbb{E} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) \right] / \exp[\varepsilon n]$ , then  $\mathbb{E} \left[ \tilde{Z}_{k,\omega,\nu}^s(\mathcal{G}(n, m)) \right] = o \left( \mathbb{E} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) \right] \right)$ .*

To prove this lemma, we combine 4.12 with [7, Corollary 1.1] and obtain the following:

**Lemma 4.17** ([7]). *Let  $s \in S_{k,\omega,\nu}$ . There is  $k_0 > 0$  such that for all  $k > k_0$  and all  $d'$  such that  $(2k-1)\ln k - 2 \leq d' \leq d_{k,\text{cond}}$  the following is true. Let  $\tau \in \tilde{\mathcal{B}}_k(n')$  be a perfectly balanced colour assignment. Then there exists  $\varepsilon > 0$  such that if  $\tilde{Z}_{k,\omega,\nu}^s(G'_{n',p',\tau})$  denotes the number of  $(\omega, n)$ -balanced  $k$ -colourings  $\tau$  of  $G'_{n',p',\tau}$  satisfying  $|\mathcal{C}(G'_{n',p',\tau}, \tau)| > \mathbb{E} \left[ Z_{k,\omega,\nu}^s(G'_{n',p',\tau}) \right] / \exp[\varepsilon n]$ , then  $\mathbb{E} \left[ \tilde{Z}_{k,\omega,\nu}^s(G'_{n',p',\tau}) \right] = o \left( \mathbb{E} \left[ Z_{k,\omega,\nu}^s(G'_{n',p',\tau}) \right] \right)$ .*

*Proof of Lemma 4.16.* Choose a map  $\sigma : [n] \rightarrow [k]$  uniformly at random and generate a graph  $G(n, p', \sigma)$  on  $[n]$  by connecting any two vertices  $v, w \in [n]$  such that  $\sigma(v) \neq \sigma(w)$  with probability  $p'$  independently. Construct  $G'_{n',p',\sigma'}$  from  $G(n, p', \sigma)$  in the way defined above. To construct  $G''_{n,p',\sigma'_{[n]}}$  from  $G'_{n',p',\sigma'}$ , we have to delete  $O(\sqrt{n})$  many vertices. By [7, Section 6], for each of these vertices  $v$  we can bound the logarithm of the number of colourings that emerge when deleting  $v$  by  $O(\ln n)$ . Thus,

$$\ln |\mathcal{C}(G''_{n,p',\sigma'_{[n]}}, \sigma'_{[n]})| = \ln |\mathcal{C}(G'_{n',p',\sigma'}, \sigma')| + O(\sqrt{n} \ln n) = \ln |\mathcal{C}(G'_{n',p',\sigma'}, \sigma')| + o(n). \quad (4.24)$$

Then Lemma 4.16 follows by combining Lemma 4.17 with (4.24) and Fact 4.13.  $\square$

To complete the proof, we have to analyse the function  $f_2$  defined in (4.4), as we know from (4.5) that

$$\mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] = \exp[nf_2(\rho) + O(\ln n)].$$

The following lemma shows that we can confine ourselves to the investigation of the function  $\bar{f}_2$  defined in (4.2).

**Lemma 4.18.** *Let  $\lim_{n \rightarrow \infty} (\rho_n)_n = \rho_0$ . Then  $\lim_{n \rightarrow \infty} \ln \mathbb{E} \left[ Z_{k,\rho_n}^{(2)}(\mathcal{G}(n, m)) \right] \leq \bar{f}_2(\rho_0)$ .*

*Proof.* Lemma 4.3 yields that

$$\exp[nf_2(\rho)] \sim \exp[n\bar{f}_2(\rho) + O(\omega^2)].$$

Together with the uniform continuity of  $\bar{f}_2$  this proves the assertion.  $\square$

We use results from [13] where an analysis of  $\bar{f}_2$  was performed. The following lemma summarizes this analysis from [13, Section 4]. The same result was used in [6].

**Lemma 4.19.** *For any  $c > 0$ , there is  $k_0 > 0$  such that for all  $k > k_0$  and all  $d$  such that  $(2k-1)\ln k - c \leq d' \leq (2k-1)\ln k$  the following statements are true.*

- (1) *If  $1 \leq s < k$ , then for all separable  $s$ -stable  $\rho \in \bar{\mathcal{B}}_k$  we have  $\bar{f}_2(\rho) < \bar{f}_2(\bar{\rho})$ .*
- (2) *If  $\rho \in \bar{\mathcal{B}}_k$  is 0-stable and  $\rho \neq \bar{\rho}$ , then  $\bar{f}_2(\rho) < \bar{f}_2(\bar{\rho})$ .*
- (3) *If  $d' = (2k-1)\ln k - 2$ , then for all separable,  $k$ -stable  $\rho \in \bar{\mathcal{B}}_k$  we have  $\bar{f}_2(\rho) < \bar{f}_2(\bar{\rho})$ .*

*Proof of Proposition 4.11.* Assume that  $k \geq k_0$  for a large enough number  $k_0$  and that  $d' \geq 2(k-1)\ln(k-1)$ . We consider two different cases.

**Case 1:**  $d' \leq (2k-1)\ln k - 2$ : Let  $\tilde{Z}_{k,\omega,\nu}^s$  be the number of  $(\omega, n)$ -balanced separable  $k$ -colourings of  $\mathcal{G}(n, m)$ .

Then Lemma 4.15 implies that  $\mathbb{E}[\tilde{Z}_{k,\omega,\nu}^s(\mathcal{G}(n, m))] \sim \mathbb{E} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) \right]$ . Furthermore, in the case that  $d' = (2k-1)\ln k - 2$ , the combination of the statements of Lemma 4.19 imply that  $\bar{f}_2(\rho) < \bar{f}_2(\bar{\rho})$  for any separable  $\rho \in \bar{\mathcal{B}}_k \setminus \{\bar{\rho}\}$ . As  $\bar{f}_2(\rho)$  is the sum of the concave function  $\rho \mapsto \mathcal{H}(\rho)$  and the convex function  $\rho \mapsto \frac{d}{2} \ln(1 - 2/k \|\rho\|_2^2)$ , this implies that, in fact, for any  $d' \leq (2k-1)\ln k - 2$  we have  $\bar{f}_2(\rho) < \bar{f}_2(\bar{\rho})$  for any separable  $\rho \in \bar{\mathcal{B}}_k \setminus \{\bar{\rho}\}$ . Hence, the uniform continuity of  $\bar{f}_2$  on  $\mathcal{B}_k$  and (4.5) yield

$$\mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))^2] \leq (1 + o(1)) \sum_{\substack{\rho \in \mathcal{B}_{k,\omega,\nu}^s(n) \\ \rho \text{ is 0-stable}}} \mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right]. \quad (4.25)$$

Additionally, as  $\bar{\mathcal{B}}_k$  is a compact set, with the second statement of Lemma 4.19 it follows that for any  $\eta > 0$  there exists  $\varepsilon > 0$  such that

$$\max_{\substack{\rho \in \mathcal{B}_{k,\omega,\nu}^s(n) \\ \rho \text{ is 0-stable} \\ \|\rho - \bar{\rho}\|_2 > \eta}} \exp[n\bar{f}_2(\rho)] \leq \exp[n\bar{f}_2(\bar{\rho}) - \varepsilon]. \quad (4.26)$$

As on the other hand it holds that

$$\mathbb{E} \left[ Z_{k,\omega,\nu,\eta}^{s(2)}(\mathcal{G}(n, m)) \right] \geq \exp[n\bar{f}_2(\bar{\rho})] / \text{poly}(n), \quad (4.27)$$

combining (4.26) and (4.27) with (4.5) and the observation that  $|\mathcal{B}_{k,\omega,\nu}^s(n)| \leq n^{k^2}$ , we see that for any  $\eta > 0$ ,

$$\sum_{\substack{\rho \in \mathcal{B}_{k,\omega,\nu}^s(n) \\ \rho \text{ is 0-stable} \\ \|\rho - \bar{\rho}\|_2 > \eta}} \mathbb{E} \left[ Z_{k,\rho}^{(2)}(\mathcal{G}(n, m)) \right] \leq \sum_{\substack{\rho \in \mathcal{B}_{k,\omega,\nu}^s(n) \\ \rho \text{ is 0-stable} \\ \|\rho - \bar{\rho}\|_2 > \eta}} \exp[n\bar{f}_2(\rho) + O(\ln n)] = o \left( \mathbb{E} \left[ Z_{k,\omega,\nu,\eta}^{s(2)}(\mathcal{G}(n, m)) \right] \right). \quad (4.28)$$

**Case 2:**  $(2k - 1) \ln k - 2 < d' < d_{k,\text{cond}}$ : For an appropriate  $\varepsilon > 0$  let  $\tilde{Z}_{k,\omega,\nu}^s$  be the number of  $(\omega, n)$ -balanced separable  $k$ -colourings  $\sigma$  of  $\mathcal{G}(n, m)$  such that  $|\mathcal{C}(\mathcal{G}(n, m), \sigma)| \leq \mathbb{E} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) \right] / \exp[\varepsilon n]$ . Then Lemmas 4.15 and 4.16 imply that  $\mathbb{E}[\tilde{Z}_{k,\omega,\nu}^s(\mathcal{G}(n, m))] \sim \mathbb{E} \left[ Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) \right]$ . Furthermore, the first part of Lemma 4.19 and equation (4.5) entail that (4.25) holds for this random variable  $\tilde{Z}_{k,\omega,\nu}^s$ . Moreover, as in the previous case (4.26), (4.27), (4.5) and the second part of Lemma 4.19 show that (4.28) holds true for any fixed  $\eta > 0$ .

In either case the assertion follows by combining (4.25) and (4.28).  $\square$

*Proof of Proposition 2.5.* The assertion is obtained by combining Proposition 2.1 with Propositions 4.11, 4.5 and 4.8.  $\square$

*Acknowledgements.* I thank my PhD advisor Amin Coja-Oghlan for constant support and valuable suggestions and Samuel Hetterich for helpful discussions.

## REFERENCES

- [1] D. Achlioptas, A. Coja-Oghlan: Algorithmic barriers from phase transitions. Proc. 49th FOCS (2008) 793–802.
- [2] D. Achlioptas, E. Friedgut: A sharp threshold for  $k$ -colorability. Random Struct. Algorithms **14** (1999) 63–70.
- [3] D. Achlioptas, A. Naor: The two possible values of the chromatic number of a random graph. Annals of Mathematics **162** (2005), pp. 1333–1349.
- [4] N. Alon, M. Krivelevich: The concentration of the chromatic number of random graphs. Combinatorica **17** (1997), pp. 303–313.
- [5] J. Banks, C. Moore: Information-theoretic thresholds for community detection in sparse networks. arXiv:1601.02658.
- [6] V. Bapst, A. Coja-Oghlan, C. Efthymiou: Planting colourings silently. arXiv:1411.0610.
- [7] V. Bapst, A. Coja-Oghlan, S. Hetterich, F. Raßmann, Dan Vilenchik: The condensation phase transition in random graph coloring. Communications in Mathematical Physics **341(2)** (2016), pp. 543–606.
- [8] V. Bapst, A. Coja-Oghlan, F. Raßmann: A positive temperature phase transition in random hypergraph 2-colouring. Annals of Applied Probability **26(3)** (2016), pp. 1362–1406.
- [9] B. Bollobás: The chromatic number of random graphs. Combinatorica **8** (1988), pp. 49–55.
- [10] B. Bollobás: Random graphs. 2nd edition. Cambridge University Press (2001).
- [11] A. Coja-Oghlan: Upper-bounding the  $k$ -colorability threshold by counting covers. Electronic Journal of Combinatorics **20** (2013), P32.
- [12] A. Coja-Oghlan, C. Efthymiou, S. Hetterich: On the chromatic number of random regular graphs. Journal of Combinatorial Theory, Series B **116**, pp. 367–439.
- [13] A. Coja-Oghlan, D. Vilenchik: Chasing the  $k$ -colorability threshold. Proc. 54th FOCS (2013), pp. 380–389. A full version is available as arXiv:1304.1063.
- [14] A. Coja-Oghlan, N. Wormald: The number of satisfying assignments of random regular  $k$ -SAT formulas. Unpublished manuscript.
- [15] P. Erdős, A. Rényi: On the evolution of random graphs. Magyar Tud. Akad. Mat. Kutató Int. Közl. **5** (1960), pp. 17–61.
- [16] S. Janson: Random regular graphs: asymptotic distributions and contiguity. Combinatorics, Probability and Computing **4** (1995), pp. 369–405.
- [17] G. Kemkes, X. Perez-Gimenez and N. Wormald: On the chromatic number of random  $d$ -regular graphs. Advances in Mathematics **223** (2010), pp. 300–328.
- [18] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborova: Gibbs states and the set of solutions of random constraint satisfaction problems. Proc. National Academy of Sciences **104** (2007), pp. 10318–10323.
- [19] T. Łuczak: The chromatic number of random graphs. Combinatorica **11** (1991), pp. 45–54.
- [20] T. Łuczak: A note on the sharp concentration of the chromatic number of random graphs. Combinatorica **11** (1991), pp. 295–297.

- [21] D. Matula: Expose-and-merge exploration and the chromatic number of a random graph. *Combinatorica* **7** (1987), pp. 275–284.
- [22] M. Molloy: The freezing threshold for  $k$ -colourings of a random graph. *Proc. 43rd STOC* (2012), pp. 921–930.
- [23] A. Montanari, R. Restrepo, P. Tetali: Reconstruction and clustering in random constraint satisfaction problems. *SIAM Journal of Discrete Mathematics* **25** (2011), pp. 771–808.
- [24] C. Moore: The phase transition in random regular exact cover. *arXiv:1502.07591*.
- [25] J. Neeman, P. Netrapalli: Non-Reconstructability in the Stochastic Block Model. Technical Report in *arXiv:1404.6304*
- [26] F. Rassmann: On the number of solutions in random hypergraph 2-colouring. *arXiv:1603.07523*.
- [27] R. Robinson, N. Wormald: Almost all cubic graphs are Hamiltonian. *Random Structures and Algorithms* **3(2)** (1992), pp. 117–125.
- [28] R. Robinson, N. Wormald: Almost all regular graphs are Hamiltonian. *Random Struct. Algorithms* **5** (1994), pp. 363–374.
- [29] E. Shamir, J. Spencer: Sharp concentration of the chromatic number of random graphs  $G(n, p)$ . *Combinatorica* **7** (1987), pp. 121–129.
- [30] N. Wormald: Models of random regular graphs. *London Mathematical Society Lecture Note Series* (1999), pp. 239–298.

## APPENDIX A.

*Proof of Corollary 2.6.* This proof is a close adaption of the analogous proof in [6]. We let  $\mathcal{E}$  denote the event  $\{\forall 2 \leq l \leq L : C_{l,n} = c_l\}$  and fix  $s \in S_{k,\omega,\nu}$ . Let  $\mathcal{Z}_n = \tilde{Z}_{k,\omega,\nu}^s(\mathcal{G}(n, m))$  for the sake of brevity. Since  $\mathcal{Z}_n \leq Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))$ , equation (2.5) yields the upper bound

$$\frac{\mathbb{E}[\mathcal{Z}_n | \mathcal{E}]}{\mathbb{E}[\mathcal{Z}_n]} \leq \frac{\mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \mathcal{E}]}{(1 + o(1))\mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))]} \sim \prod_{l=2}^L [1 + \delta_l]^{c_l} \exp[-\delta_l \lambda_l]. \quad (\text{A.1})$$

We show the following matching lower bound:

$$\mathbb{E}[\mathcal{Z}_n | \mathcal{E}] \geq (1 - o(1))\mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \mathcal{E}]. \quad (\text{A.2})$$

Indeed, assume for contradiction that (A.2) is false. Then we can find an  $n$ -independent  $\varepsilon > 0$  such that for infinitely many  $n$ ,

$$\mathbb{E}[\mathcal{Z}_n | \mathcal{E}] < (1 - \varepsilon)\mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \mathcal{E}]. \quad (\text{A.3})$$

By Fact 2.2 there exists an  $n$ -independent  $\xi = \xi(c_2, \dots, c_L) > 0$  such that  $\mathbb{P}[\mathcal{E}] \geq \xi$ . Hence, (A.3) and Bayes' formula imply that

$$\begin{aligned} \mathbb{E}[\mathcal{Z}_n] &= \mathbb{E}[\mathcal{Z}_n | \mathcal{E}] \mathbb{P}[\mathcal{E}] + \mathbb{E}[\mathcal{Z}_n | \neg \mathcal{E}] \mathbb{P}[\neg \mathcal{E}] \\ &\leq (1 - \varepsilon)\mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \mathcal{E}] \mathbb{P}[\mathcal{E}] + \mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \neg \mathcal{E}] \mathbb{P}[\neg \mathcal{E}] \\ &\leq \mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))] - \varepsilon \xi \cdot \mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | \mathcal{E}] \\ &= \mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))] \cdot \left(1 + o(1) - \varepsilon \xi \prod_{l=2}^L [1 + \delta_l]^{c_l} \exp[-\delta_l \lambda_l]\right) \\ &= (1 - \Omega(1))\mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))], \end{aligned} \quad (\text{A.4})$$

where the last equality holds since  $\delta_l, \lambda_l$  and  $c_l$  remain fixed as  $n \rightarrow \infty$ . As (A.4) contradicts (2.5), we have established (A.2). Finally, combining (A.2) with (2.4) and (2.5), we get

$$\frac{\mathbb{E}[\mathcal{Z}_n | S]}{\mathbb{E}[\mathcal{Z}_n]} \geq \frac{(1 - o(1))\mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m)) | S]}{(1 + o(1))\mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))]} \sim \prod_{l=2}^L [1 + \delta_l]^{c_l} \exp[-\delta_l \lambda_l], \quad (\text{A.5})$$

and the assertion follows from (A.1) and (A.5). □

**A.1. Calculating the first moment.** The following proofs are very close to analogous proofs in [6].

*Proof of Lemma 3.1.* As the edges in  $\mathcal{G}(n, m)$  are independent by construction, the expected number of  $k$ -colourings with colour density  $\rho$  is given by

$$\mathbb{E}[Z_{k,\rho}(\mathcal{G}(n, m))] = \binom{n}{\rho_1 n, \dots, \rho_k n} \left(1 - \frac{1}{N} \sum_{i=1}^k \binom{\rho_i n}{2}\right)^m, \quad \text{where } N = \binom{n}{2}. \quad (\text{A.6})$$

Further, the number of forbidden edges is given by

$$\sum_{i=1}^k \binom{\rho_i n}{2} = N \left( \sum_{i=1}^k \rho_i^2 \right) + \frac{n}{2} \left( \sum_{i=1}^k \rho_i^2 - 1 \right) + O(1)$$

and thus

$$\begin{aligned} m \ln \left( 1 - \frac{1}{N} \sum_{i=1}^k \binom{\rho_i n}{2} \right) &= m \ln \left[ \left( 1 + \frac{n}{2N} \right) \left( 1 - \sum_{i=1}^k \rho_i^2 \right) \right] + o(1) \\ &= n \frac{d}{2} \ln \left( 1 - \sum_{i=1}^k \rho_i^2 \right) + \frac{d}{2} + o(1). \end{aligned} \quad (\text{A.7})$$

Equation (3.1) follows from (A.6), (A.7) and Stirling's formula. Moreover, (3.2) follows from (A.6) and (A.7) because  $\|\rho - \rho^*\|_2 = o(1)$  implies that  $\sum_{i=1}^k \rho_i^2 \sim 1/k$  and

$$\binom{n}{\rho_1 n, \dots, \rho_k n} \sim (2\pi n)^{\frac{1-k}{2}} k^{k/2} \exp[n\mathcal{H}(\rho)].$$

□

*Proof of Corollary 3.2.* The functions  $\rho \in \mathcal{A}_k \mapsto \mathcal{H}(\rho)$  and  $\rho \in \mathcal{A}_k \mapsto \frac{d}{2} \ln(1 - \sum_{i=1}^k \rho_i^2)$  are both concave and attain their maximum at  $\rho = \rho^*$ . Consequently, setting  $B(d, k) = k(1 + \frac{d}{k-1})$  and expanding around  $\rho = \rho^*$ , we obtain

$$f_1(\rho^*) - \frac{B(d, k)}{2} \|\rho - \rho^*\|_2^2 - O(\|\rho - \rho^*\|_2^3) \leq f_1(\rho) \leq f_1(\rho^*) - \frac{B(d, k)}{2} \|\rho - \rho^*\|_2^2. \quad (\text{A.8})$$

Plugging the upper bound from (A.8) into (3.1) and observing that  $|\mathcal{A}_k(n)| \leq n^k = \exp[o(n)]$ , we find

$$S_1 = \sum_{\substack{\rho \in \mathcal{A}_k(n) \\ \|\rho - \rho^*\|_2 > n^{-3/8}}} \mathbb{E}[Z_{k, \rho}(\mathcal{G}(n, m))] \leq C_2 \exp[f_1(\rho^*)] \exp\left[-\frac{B(d, k)}{2} n^{1/6}\right]. \quad (\text{A.9})$$

On the other hand, (3.2) implies that

$$\begin{aligned} S_2 &= \sum_{\substack{\rho \in \mathcal{A}_k(n) \\ \|\rho - \rho^*\|_2 \leq n^{-3/8}}} \mathbb{E}[Z_{k, \rho}(\mathcal{G}(n, m))] \sim \sum_{\substack{\rho \in \mathcal{A}_k(n) \\ \|\rho - \rho^*\|_2 \leq n^{-3/8}}} (2\pi n)^{\frac{1-k}{2}} k^{k/2} \exp[d/2] \exp[nf_1(\rho)] \\ &\sim (2\pi n)^{\frac{1-k}{2}} k^{k/2} \exp[d/2 + nf_1(\rho^*)] \sum_{\rho \in \mathcal{A}_k(n)} \exp\left[-n \frac{B(d, k)}{2} \|\rho - \rho^*\|_2^2\right]. \end{aligned} \quad (\text{A.10})$$

The last sum is nearly in the standard form of a Gaussian summation, merely the vectors  $\rho \in \mathcal{A}_k(n)$  that we sum over are subject to the linear constraint  $\rho_1 + \dots + \rho_k = 1$ . We rid ourselves of this constraint by substituting  $\rho_k = 1 - \rho_1 - \dots - \rho_{k-1}$ . Formally, let  $J$  be the  $(k-1) \times (k-1)$ -matrix with diagonal entries equal to 2 and remaining entries equal to 1. We observe that  $\det J = k$ . Then

$$\begin{aligned} \sum_{\rho \in \mathcal{A}_k(n)} \exp\left[-n \frac{B(d, k)}{2} \|\rho - \rho^*\|_2^2\right] &\sim \sum_{y \in \frac{1}{n}\mathbb{Z}^k} \exp\left[-n \frac{B(d, k)}{2} \langle Jy, y \rangle\right] \\ &\sim (2\pi n)^{\frac{k-1}{2}} k^{-\frac{k}{2}} \left(1 + \frac{d}{k-1}\right)^{-\frac{k-1}{2}}. \end{aligned} \quad (\text{A.11})$$

Plugging (A.11) into (A.10), we obtain

$$\begin{aligned} S_2 &\sim (2\pi n)^{\frac{1-k}{2}} k^{k/2} \exp[d/2 + nf_1(\rho^*)] (2\pi n)^{\frac{k-1}{2}} k^{-\frac{k}{2}} \left(1 + \frac{d}{k-1}\right)^{-\frac{k-1}{2}} \\ &= \exp[d/2 + nf_1(\rho^*)] \left(1 + \frac{d}{k-1}\right)^{-\frac{k-1}{2}}. \end{aligned} \quad (\text{A.12})$$



Finally, comparing (A.9) and (A.12), we see that  $S_1 = o(S_2)$ . Thus,  $\mathbb{E}[Z_k(\mathcal{G}(n, m))] = S_1 + S_2 \sim S_2$ , and the assertion follows from (A.12).  $\square$

**A.2. Calculating the second moment.** The following proofs are very close to analogous proofs in [6].

*Proof of 4.1.* To calculate the expected number of pairs of colourings  $\sigma, \tau$  with overlap  $\rho \in \mathcal{B}_k(n)$ , we first observe that

$$\mathbb{P}[\sigma, \tau \text{ are } k\text{-colourings of } \mathcal{G}(n, m)] = \left(1 - \frac{\mathcal{F}(\sigma, \tau)}{N}\right)^m,$$

where  $\mathcal{F}(\sigma, \tau)$  is the number of “forbidden” edges joining two vertices with the same colour under either  $\sigma$  or  $\tau$  and  $N = \binom{n}{2}$ . We have

$$\begin{aligned} \mathcal{F}(\sigma, \tau) &= \sum_{i=1}^k \binom{\rho_{i*}n}{2} + \sum_{j=1}^k \binom{\rho_{*j}n}{2} - \sum_{i,j=1}^k \binom{\rho_{ij}n}{2} \\ &= N \left( \sum_{i=1}^k \rho_{i*}^2 + \sum_{j=1}^k \rho_{*j}^2 - \sum_{i,j=1}^k \rho_{ij}^2 \right) + \frac{n}{2} \left( \sum_{i=1}^k \rho_{i*}^2 + \sum_{j=1}^k \rho_{*j}^2 - \sum_{i,j=1}^k \rho_{ij}^2 - 1 \right) + O(1) \end{aligned}$$

and thus, the probability that  $\sigma$  and  $\tau$  are both colourings of  $\mathcal{G}(n, m)$  only depends on their overlap  $\rho$  and is given by

$$\mathbb{P}[\sigma, \tau \text{ are } k\text{-colourings of } \mathcal{G}(n, m)] \sim \exp \left[ m \ln \left( 1 - \sum_{i=1}^k \rho_{i*}^2 - \sum_{j=1}^k \rho_{*j}^2 + \sum_{i,j=1}^k \rho_{ij}^2 \right) + \frac{d}{2} \right]. \quad (\text{A.13})$$

It remains to multiply this by the total number of  $\sigma, \tau$  with overlap  $\rho \in \mathcal{B}_k(n)$ . By Stirling’s formula, this number is given by

$$\binom{n}{\rho_{11}n, \dots, \rho_{kk}n} \sim \sqrt{2\pi n}^{-\frac{k^2-1}{2}} \left( \prod_{i,j} \frac{1}{\sqrt{2\pi\rho_{ij}}} \right) \exp[n\mathcal{H}(\rho)]. \quad (\text{A.14})$$

Equation (4.2) is obtained by combining (A.13) and (A.14). To prove (4.3), we observe that if  $\|\rho - \bar{\rho}\|_2^2 = o(1)$ , we have

$$\frac{\sqrt{2\pi n}^{\frac{1-k^2}{2}}}{\prod_{i,j=1}^k \sqrt{2\pi\rho_{ij}}} \sim k^{k^2} (2\pi n)^{\frac{1-k^2}{2}}$$

and the statement follows.  $\square$

*Proof of Lemma 4.9.* Together with the Euler-Maclaurin formula and Lemma 4.10, a Gaussian integration yields

$$\begin{aligned} \sum_{\epsilon \in S_n} \exp \left[ -n \frac{D}{2} \|\epsilon\|_2^2 + o(n^{1/2}) \|\epsilon\|_2 \right] &= \sum_{\epsilon \in (\mathbb{Z}/n)^{(k-1)^2}} \exp \left[ -n \frac{D}{2} \sum_{i,j,i',j' \in [k-1]} \mathcal{H}_{(i,j),(i',j')} \epsilon_{ij} \epsilon_{i'j'} + o(n^{1/2}) \|\epsilon\|_2 \right] \\ &\sim n^{(k-1)^2} \int \dots \int \exp \left[ -n \frac{D}{2} \sum_{i,j,i',j' \in [k-1]} \mathcal{H}_{(i,j),(i',j')} \epsilon_{ij} \epsilon_{i'j'} \right] d\epsilon_{11} \dots d\epsilon_{(k-1)(k-1)} \\ &\sim \left( \sqrt{2\pi n} \right)^{(k-1)^2} D^{-\frac{(k-1)^2}{2}} (\det \mathcal{H})^{-1/2} \sim \left( \sqrt{2\pi n} \right)^{(k-1)^2} D^{-\frac{(k-1)^2}{2}} k^{-(k-1)}, \end{aligned}$$

as desired.  $\square$

**A.3. Counting short cycles.** In this section we count the number of cycles of a short fixed length in order to prove Proposition 2.3. The results in this section were already obtained in [6] and the proofs are a very close adaption of the ones in [6]. We recall that for  $l = 2, \dots, L$  we denoted by  $C_{l,n}$  the number of cycles of length exactly  $l$  in  $\mathcal{G}(n, m)$ . We let  $c_2, \dots, c_L$  be a sequence of non-negative integers and  $\mathcal{E}$  the event that  $C_{l,n} = c_l$  for  $l = 2, \dots, L$ . We recall  $\lambda_l, \delta_l$  from (2.3). For a map  $\sigma : [n] \mapsto [k]$ , we define  $\mathcal{V}(\sigma)$  as the event that  $\sigma$  is a  $k$ -colouring of the random graph  $\mathcal{G}(n, m)$ . Our starting point is the following lemma concerning the distribution of the random variables  $C_{l,n}$  given  $\mathcal{V}(\sigma)$ .

**Lemma A.1.** *Let  $\mu_l = \frac{d^l}{2l} \left[ 1 + \frac{(-1)^l}{(k-1)^{l-1}} \right]$ . Then  $\mathbb{P}[\mathcal{E}|\mathcal{V}(\sigma)] \sim \prod_{l=2}^L \frac{\exp[-\mu_l]}{c_l!} \mu_l^{c_l}$  for any  $\sigma$  with  $\rho(\sigma) \in \mathcal{A}_{k,\omega}(n)$ .*

*Proof.* All we have to show is that for any fixed sequence of integers  $m_2, \dots, m_L \geq 0$ , the joint factorial moments satisfy

$$\mathbb{E}[(C_{2,n})_{m_2} \cdots (C_{L,n})_{m_L} | \mathcal{V}(\sigma)] \sim \prod_{l=2}^L \mu_l^{m_l}. \quad (\text{A.15})$$

Then Lemma A.1 follows from [10, Theorem 1.23].

To establish (A.15), we interpret  $(C_{2,n})_{m_2} \cdots (C_{L,n})_{m_L}$  as the number of sequences of  $m_2 + \cdots + m_L$  distinct cycles such that  $m_2$  is the number of cycles of length 2, and so on. We let  $Y$  be the number of those sequences of cycles such that any two cycles are vertex-disjoint and  $Y'$  be the number of sequences having intersecting cycles. Obviously, we have

$$\mathbb{E}[(C_{2,n})_{m_2} \cdots (C_{L,n})_{m_L} | \mathcal{V}(\sigma)] = \mathbb{E}[Y | \mathcal{V}(\sigma)] + \mathbb{E}[Y' | \mathcal{V}(\sigma)]. \quad (\text{A.16})$$

For  $\mathbb{E}[Y' | \mathcal{V}(\sigma)]$ , we use the following claim that we prove at the end of this section.

**Claim A.2.** *It holds that  $\mathbb{E}[Y' | \mathcal{V}(\sigma)] = O(n^{-1})$ .*

Thus, it remains to count the number of vertex disjoint cycles conditioned on  $\mathcal{V}(\sigma)$ . The line of arguments we use is similar to [17, Section 2]. To simplify the calculations, we define  $D_{l,n}$  as the number of rooted, directed cycles of length  $l$  in  $\mathcal{G}(n, m)$ , implying that  $D_{l,n} = 2lC_{l,n}$ .

For a rooted directed cycle  $(v_1, \dots, v_l)$  of length  $l$ , we call  $(\sigma(v_1), \dots, \sigma(v_l))$  the *type* of the cycle under  $\sigma$ . Let  $D_{l,n}^t$  denote the number of rooted, directed cycles of length  $l$  and type  $t = (t_1, \dots, t_l)$ . We claim that

$$\mathbb{E}[D_{l,n}^t | \mathcal{V}(\sigma)] \sim \left(\frac{n}{k}\right)^l \frac{(m)_l}{(N - \mathcal{F}(\sigma))^l} \sim \left(\frac{d}{k-1}\right)^l \quad \text{with } N = \binom{n}{2}. \quad (\text{A.17})$$

Indeed, as  $\sigma$  is  $(\omega, n)$ -balanced, the number of ways of choosing  $l$  vertices  $(v_1, \dots, v_l)$  such that  $\sigma(v_i) = t_i$  for all  $i$  is  $(1 + o(1))(n/k)^l$  and each edge  $\{v_i, v_{i+1}\}$  of the cycle is present in the graph with a probability asymptotically equal to  $m/(N - \mathcal{F}(\sigma))$ . This explains the first asymptotic equality in (A.17). The second one follows because  $m = dn/2$  and  $\mathcal{F}(\sigma) \sim N/k$ .

In particular, the r.h.s. of (A.17) is independent of the type  $t$ . For a given  $l$ , let  $T_l$  signify the number of all possible types of cycles of length  $l$ . Thus,  $T_l$  is the set of all sequences  $(t_1, \dots, t_l)$  such that  $t_{i+1} \neq t_i$  for all  $1 \leq i < l$  and  $t_l \neq t_1$ . Let  $T_1 = 0$ . Then  $T_l$  satisfies the recurrence

$$T_l + T_{l-1} = k(k-1)^{l-1}. \quad (\text{A.18})$$

To see this, observe that  $k(k-1)^{l-1}$  is the number of all sequences  $(t_1, \dots, t_l)$  such that  $t_{i+1} \neq t_i$  for all  $1 \leq i < l$ . Any such sequence either satisfies  $t_l \neq t_1$ , which is accounted for by  $T_l$ , or  $t_l = t_1$  and  $t_{l-1} \neq t_1$ , in which case it is contained in  $T_{l-1}$ .

Hence, iterating (A.18) gives  $T_l = (k-1)^l + (-1)^l(k-1)$ . Combining this formula with (A.17), we obtain

$$\mathbb{E}[D_{l,n} | \mathcal{V}(\sigma)] \sim T_l \cdot \mathbb{E}(D_{l,n}^t | \mathcal{V}(\sigma)) \sim d^l \left( 1 + \frac{(-1)^l}{(k-1)^{l-1}} \right).$$

Recalling that  $C_{l,n} = D_{l,n}/(2l)$ , we get

$$\mathbb{E}[C_{l,n} | \mathcal{V}(\sigma)] \sim \frac{d^l}{2l} \left( 1 + \frac{(-1)^l}{(k-1)^{l-1}} \right). \quad (\text{A.19})$$

Since  $Y$  considers only vertex disjoint cycles and  $l, m_2, \dots, m_L$  remain fixed as  $n \rightarrow \infty$ , equation (A.19) yields

$$\mathbb{E}[Y|\mathcal{V}(\sigma)] \sim \prod_{l=2}^L \left( \frac{d^l}{2l} \left( 1 + \frac{(-1)^l}{(k-1)^{l-1}} \right) \right)^{m_l}.$$

Plugging the above relation and Claim A.2 into (A.16), we get (A.15) and the assertion follows.  $\square$

**Proof of Proposition 2.3:** Let  $s \in S_{k,\omega,\nu}$ . By Bayes' rule and Lemma A.1 we have

$$\begin{aligned} \mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))|\mathcal{E}] &= \frac{1}{\mathbb{P}[\mathcal{E}]} \sum_{\tau \in \mathcal{A}_{k,\omega,\nu}^s(n)} \mathbb{P}[\mathcal{V}(\tau)] \mathbb{P}[\mathcal{E}|\mathcal{V}(\tau)] \\ &\sim \frac{\prod_{l=2}^L \frac{\exp[-\mu_l]}{c_l!} \mu_l^{c_l}}{\mathbb{P}[\mathcal{E}]} \sum_{\tau \in \mathcal{A}_{k,\omega,\nu}^s(n)} \mathbb{P}[\mathcal{V}(\tau)] \\ &\sim \frac{\prod_{l=2}^L \frac{\exp[-\mu_l]}{c_l!} \mu_l^{c_l}}{\mathbb{P}[\mathcal{E}]} \mathbb{E}[Z_{k,\omega,\nu}^s(\mathcal{G}(n, m))]. \end{aligned}$$

From Lemma A.1 and Fact 2.2 we get that

$$\frac{\prod_{l=2}^L \frac{\exp[-\mu_l]}{c_l!} \mu_l^{c_l}}{\mathbb{P}[\mathcal{E}]} \sim \prod_{l=2}^L [1 + \delta_l]^{c_l} \exp[-\delta_l \lambda_l],$$

whence Proposition 2.3 follows.  $\square$

**Proof of Claim A.2:** For every subset  $R$  of  $l \leq L$  vertices, let  $\mathbb{I}_R$  be equal to 1 if the number of edges with both ends in  $R$  is at least  $|R| + 1$ . Let  $H_L$  be the event that  $\{\sum_{R: |R| \leq L} \mathbb{I}_R > 0\}$ . By definition, if  $Y' > 0$  then the event  $H_L$  occurs. This implies that

$$\mathbb{P}[Y' > 0|\mathcal{V}(\sigma)] \leq \mathbb{P}[H_L|\mathcal{V}(\sigma)].$$

Thus, it suffices to appropriately bound  $\mathbb{P}[H_L|\mathcal{V}(\sigma)]$ . Markov's inequality yields

$$\mathbb{P}[H_L|\mathcal{V}(\sigma)] \leq \mathbb{E} \left[ \sum_{R: |R| \leq L} \mathbb{I}_R | \mathcal{V}(\sigma) \right] = \sum_{l=2}^L \sum_{R: |R|=l} \mathbb{E}[\mathbb{I}_R | \mathcal{V}(\sigma)].$$

For any set  $R$  such that  $|R| = l$ , we can put  $l + 1$  edges inside the set in at most  $\binom{l}{l+1}$  ways. Clearly conditioning on  $\mathcal{V}(\sigma)$  can only reduce the number of different placings of the edges. For a fixed set  $R$  of cardinality  $l$ , we get, using inclusion/exclusion and the Binomial theorem as well as the fact that  $\mathcal{F}(\sigma) \sim N/k$ :

$$\begin{aligned} \mathbb{E}[\mathbb{I}_R | \mathcal{V}(\sigma)] &\leq \binom{l}{l+1} \sum_{i=0}^{l+1} \binom{l+1}{i} (-1)^i \left( 1 - \frac{i}{N - \mathcal{F}(\sigma)} \right)^m \\ &\leq \binom{l}{l+1} \left( \frac{m}{N - \mathcal{F}(\sigma)} \right)^{l+1} \sim \binom{l}{l+1} \left( \frac{d}{n(1-1/k)} \right)^{l+1}. \end{aligned}$$

As  $\binom{i}{j} \leq (ie/j)^j$ , it follows that

$$\begin{aligned} \mathbb{P}[H_L|\mathcal{V}(\sigma)] &\leq (1 + o(1)) \sum_{l=2}^L \binom{n}{l} \binom{l}{l+1} \left( \frac{d}{n(1-1/k)} \right)^{l+1} \\ &\leq (1 + o(1)) \sum_{l=2}^L \left( \frac{ne}{l} \right)^l \left( \frac{le}{2} \right)^{l+1} \left( \frac{d}{n(1-1/k)} \right)^{l+1} \\ &\leq \frac{1 + o(1)}{n} \sum_{l=2}^L \frac{led}{2(1-1/k)} \left( \frac{e^2 d}{2(1-1/k)} \right)^l = O(n^{-1}), \end{aligned}$$

where the last equality holds since  $L$  is a fixed number. This proves the claim.  $\square$

FELICIA RASSMANN, [rassmann@math.uni-frankfurt.de](mailto:rassmann@math.uni-frankfurt.de), GOETHE UNIVERSITY, INSTITUTE OF MATHEMATICS, 10 ROBERT-MAYER-STR, FRANKFURT 60325, GERMANY.